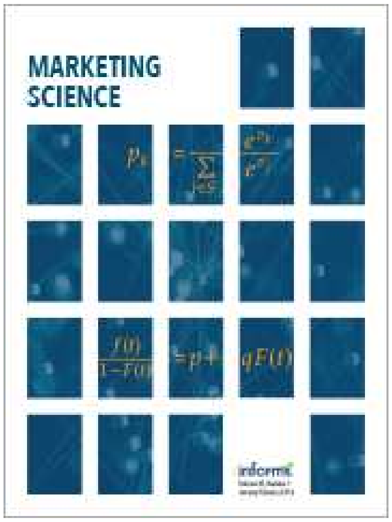
The role of time-varying contextual factors in latent attrition models for customer base analysis

Bachmann, Patrick ; Meierer, Markus ; Näf, Jeffrey

Abstract: Customer base analysis of noncontractual businesses builds on modeling purchases and latent attrition. With the Pareto/NBD model, this has become a straightforward exercise. However, this simplicity comes at a price. Customer-level predictions often lack precision. This issue can be addressed by acknowledging the importance of contextual factors for customer behavior. Considering contextual factors might contribute in two ways: (1) by increasing predictive accuracy and (2) by identifying the impact of these determinants on the purchase and attrition process. However, there is no generalization of the Pareto/NBD model that incorporates time-varying contextual factors. Preserving a closed-form maximum likelihood solution, this study proposes an extension that facilitates modeling time-invariant and time-varying contextual factors in continuous noncontractual settings. These contextual factors can influence the purchase process, the attrition process, or both. The authors further illustrate how to control for endogenous contextual factors. Benchmarking with three data sets from the retailing industry shows that explicitly modeling time-varying contextual factors significantly improves the accuracy of out-of-sample predictions for future purchases and latent attrition.

DOI: https://doi.org/10.1287/mksc.2020.1254

Originally published at:
Bachmann, Patrick; Meierer, Markus; Näf, Jeffrey (2021). The role of time-varying contextual factors in latent attrition models for customer base analysis. Marketing Science, 40(4):783-809.
DOI: https://doi.org/10.1287/mksc.2020.1254



The Role of Time-Varying Contextual Factors in Latent Attrition Models for Customer Base Analysis

Patrick Bachmann, Markus Meierer, Jeffrey Näf

informs
http://pubsonline.informs.org/journal/mksc

MARKETING SCIENCE
Articles in Advance, pp. 1–38
ISSN 0732-2399 (print), ISSN 1526-548X (online)

The Role of Time-Varying Contextual Factors in Latent Attrition Models for Customer Base Analysis

Patrick Bachmann,[a] Markus Meierer,[a] Jeffrey Näf[b]

[a] University Research Priority Program Social Networks, University of Zurich, 8050 Zurich, Switzerland; [b] Department of Mathematics, Eidgenössische Technische Hochschule (ETH) Zurich, 8092 Zurich, Switzerland

Contact: patrick.bachmann@business.uzh.ch, https://orcid.org/0000-0002-3938-2697 (PB); markus.meierer@business.uzh.ch, https://orcid.org/0000-0003-1995-5897 (MM); naef@stat.math.ethz.ch, https://orcid.org/0000-0003-0920-1899 (JN)

Received: July 5, 2017
Revised: August 3, 2018; April 16, 2019; September 17, 2019; March 12, 2020
Accepted: May 12, 2020
Published Online in Articles in Advance: January 15, 2021

https://doi.org/10.1287/mksc.2020.1254

Abstract. Customer base analysis of noncontractual businesses builds on modeling purchases and latent attrition. With the Pareto/NBD model, this has become a straightforward exercise. However, this simplicity comes at a price. Customer-level predictions often lack precision. This issue can be addressed by acknowledging the importance of contextual factors for customer behavior. Considering contextual factors might contribute in two ways: (1) by increasing predictive accuracy and (2) by identifying the impact of these determinants on the purchase and attrition process. However, there is no generalization of the Pareto/NBD model that incorporates time-varying contextual factors. Preserving a closed-form maximum likelihood solution, this study proposes an extension that facilitates modeling time-invariant and time-varying contextual factors in continuous noncontractual settings. These contextual factors can influence the purchase process, the attrition process, or both. The authors further illustrate how to control for endogenous contextual factors. Benchmarking with three data sets from the retailing industry shows that explicitly modeling time-varying contextual factors significantly improves the accuracy of out-of-sample predictions for future purchases and latent attrition.

History: Yuxin Chen served as the senior editor and Peter Fader served as associate editor for this article.

Funding: This work was funded by the University Research Priority Program Social Networks at the University of Zurich and the Swiss National Science Foundation [Project 100018_163189].
Supplemental Material: The data files and online appendix are available at https://doi.org/10.1287/mksc.2020.1254.

Keywords: probability models • Pareto/NBD • customer relationship management • latent attrition • contextual factors • customer lifetime value

1. Introduction

Modeling customer purchases and attrition in noncontractual businesses has become a straightforward task, but this simplicity comes at a price. Having access to recency and frequency data of customers' past transactions allows marketers to apply the Pareto/NBD model (Schmittlein et al. 1987). However, predictions are said to represent an educated guess rather than a precise value (Wübben and Wangenheim 2008, Malthouse 2009, Fader 2012). If the focus is on the aggregated level, that is, the entire customer base, this difference can be rather negligible. In contrast, the applicability of individual level predictions is often limited. For example, an increased level of precision is required when allocating resources for customer retention activities to individual customers.

There have been many attempts to improve the original Pareto/NBD model. However, there exists no generalization that allows modeling time-varying contextual factors in a continuous noncontractual setting. Extensions to the Pareto/NBD model or related models mainly focus on the computational complexity of the estimation procedure (Fader et al. 2005a), the correlation between the modeled processes (Glady et al. 2015), or the integration of time-invariant contextual factors (Fader and Hardie 2007, Abe 2009, Singh et al. 2009). Recent studies illustrate how regularity patterns (Platzer and Reutterer 2016) and stationary transaction attributes can be included in the customer's purchase process (Braun et al. 2015) or how differences across customer cohorts may be captured in latent attrition models (Gopalakrishnan et al. 2017). An approach by Schweidel and Knox (2013) goes one step further and provides the possibility of including other time-varying contextual factors (i.e. direct mailing activity) into a discrete latent attrition model. Although the Pareto/NBD model is very popular in research and practice, no extension allows for the inclusion of time-varying contextual factors.

The impact of two broad categories of time-varying contextual factors on customer behavior has been highlighted in the previous literature (Schweidel and Knox 2013, Hanssens and Pauwels 2016): (1) seasonality in purchase patterns and (2) tactical marketing activities. Seasonality in purchase patterns is common in many noncontractual settings, especially in the retailing industry. Customer behavior is heavily influenced at the aggregate level by public holidays (e.g., Christmas and Thanksgiving) and at the individual level by recurring personal events (e.g., birthdays and paydays). In addition to seasonality, tactical marketing activities are an important time-varying determinant of customer behavior. Customers can be targeted by either individual- or aggregate-level tactical marketing activities, such as personalized or mass marketing campaigns. Explicitly modeling marketing variables, such as time-varying contextual factors, explains the heterogeneity across customers that was introduced by the firm in the first place. Including either kind of time-varying contextual factors improves the predictive accuracy of probabilistic customer attrition models. However, the impact of these contextual factors is likely to vary. Probabilistic customer attrition models operationalize customer behavior through two processes: (1) the purchase process and (2) the attrition process. When contextual factors are added to the model, they are allowed to influence customer behavior through these two processes. However, these factors do not necessarily have to affect both processes. It is reasonable that some of the contextual factors influence only customers' purchases or attrition. For example, although personalized couponing will likely impact only the purchase process, direct mailing information about loyalty program events might only impact the attrition process. Furthermore, it is a valid assumption that marketing activities irritate customers and thus increase customer attrition (Ascarza et al. 2016). Similarly, seasonality patterns, such as those caused by holiday seasons, are likely to have an exclusive impact on the purchase process, whereas other seasonality patterns, such as those caused by individual paydays might additionally impact customer attrition. The latter may cause customers to reconsider their existing business relationships. Being able to test such hypotheses is a desired feature when analyzing contextual factors in latent attrition models.

In this paper, we propose a latent attrition model that allows time-varying contextual factors to be modeled in continuous noncontractual settings. Complementing previous literature, we combine the following characteristics in the proposed approach: (1) the continuous nature of both the purchase and the attrition processes; (2) the inclusion of multiple time-varying and time-invariant contextual factors that can separately influence both, one, or none of the processes; (3) gamma heterogeneity for both processes; (4) the ability to reduce to the Standard Pareto/NBD model when it is estimated without any contextual factors; (5) a closed-form maximum-likelihood solution; and (6) the derivation of relevant managerial expressions. We conduct two simulation studies and an empirical analysis of three retailing datasets. Benchmarking the proposed approach against state-of-the-art Pareto- and non–Pareto-type models, the results provide evidence on the inferential and predictive ability of the extended Pareto/NBD model. We find that (a) predictive accuracy generally increases when the contextual factors are included and (b) differences in the increase in predictive accuracy depend on the scope of the modeled contextual factors, that is, individual level contextual factors increase predictive accuracy more than aggregated-level contextual factors. Furthermore, we can (c) reliably identify the impact of the exogenous factors on both, the purchase and the attrition process and (d) reliably identify the impact of endogenous factors on both processes when applying (latent) instrumental variable (IV) approaches. Last, we provide evidence that (e) controlling for endogeneity has a key role in reliably identifying parameter estimates but little importance for predictive accuracy. A latent attrition model with such characteristics and performance could be used for numerous managerial applications. For example, combined with a Gamma/Gamma model (Colombo and Jiang 1999, Fader et al. 2005b), this model enables academics and managers to improve the identification of the best future customers. In addition, controlling for endogenous contextual factors allows for the rigorous identification and quantification of drivers of customers' purchase and attrition processes.

The remainder of the paper is structured as follows. In the next section, we provide an overview of the relevant literature on latent customer attrition models. Then, we present our modeling framework. We derive the likelihood function and related expressions for predicting latent customer attrition and discuss model identification. In this context, we then propose three approaches for addressing potential endogeneity of the contextual factors. Next, the model is empirically validated with three real-world datasets from the retailing industry. We compare the proposed model against the standard Pareto/NBD model, three related Pareto-type models (Schweidel and Knox 2013, Braun et al. 2015, Platzer and Reutterer 2016), and a recently published model, which builds on an Bayesian nonparametric framework with Gaussian process priors (Dew and Ansari 2018). We conclude with a discussion of limitations and future research possibilities.

2. Related Research

In this section, we present an overview of the Pareto/NBD model and its extensions. We start with introducing the standard Pareto/NBD model and the related BG/NBD model. Next, we discuss extensions to include contextual factors, followed by advancements that focus on modifying the underlying processes. Finally, we extend our discussion to review the latest approaches based on non–Pareto-type models. Table 1 provides an overview of this research.

Schmittlein et al. (1987) propose a Pareto/NBD model that is capable of simultaneously modeling customers' lifetime and transaction behavior. Previously, researchers focused on modeling only transaction behavior by relying on NBD-type models (Ehrenberg 1959). By additionally deriving the probability that an individual customer is *alive* (commonly defined as P(alive)), the Pareto/NBD model addresses one of the primary challenges in measuring customer behavior in noncontractual settings (Singh et al. 2009). The probability that a customer is alive has evolved into a key metric for assessing customer lifetime value in noncontractual settings (Schmittlein and Peterson 1994; Reinartz and Kumar 2000, 2003). In subsequent research, Fader et al. (2005a) present the BG/NBD model, which performs comparable to the Pareto/NBD model but with lower computational complexity. In contrast to the Pareto/NBD model, the BG/NBD model restricts customer attrition to a discrete process and therefore allows customer churn only during repurchase incidents. Both models have been instrumental in promoting the application of customer lifetime value (CLV). However, they have a serious limitation: contextual factors such as customer characteristics, marketing activities, or seasonality patterns are neglected.

Various authors address this limitation by introducing extensions to include time-invariant contextual factors. Fader and Hardie (2007) discuss time-invariant contextual factors for the BG/NBD and Pareto/NBD models. Abe (2009) uses a Markov Chain Monte Carlo (MCMC)-based approach to account for time-invariant contextual factors, such as the initial purchase amount and demographics. Likewise, Singh et al. (2009) propose a MCMC-based data augmentation framework to consider time-invariant contextual

Table 1. Related Research

Study	Setting	Time-invariant factors: Attrition	Time-invariant factors: Purchase	Time-varying factors: Attrition	Time-varying factors: Purchase	Endogeneity control	Maximum likelihood	Heterogeneity	DERT/ DECT	Latent attrition
Schmittlein, Morrison, and Colombo (1987)	cont						x	dist		x
Gupta (1991)	cont				x		x	dist		
Fader and Hardie (2007)	cont	x	x				x	dist		x
Abe (2009)	cont	x	x					dist		x
Singh, Borle, and Jain (2009)	cont	x	x					dist		x
Neslin and Rhoads (2009)	cont	x	x					dist		x
Van Oest and Knox (2011)	cont			x[a]	x[a]		x	lc		x
Schweidel and Knox (2013)	disc	x	x	x	x	x	x	lc		x
Knox and van Oest (2014)	cont			x[a]	x[a]	x	x	lc		x
Braun and Schweidel, and Stein (2015)	cont	x		x			x	dist	x	x
Wünderlich (2015)	cont			x	x			dist		x
Harman (2016)	cont		x		x			dist		x
Platzer and Reutterer (2016)	cont	x	x					dist		x
Gopalakrishnan, Bradlow and Fader (2017)	cont	x	x	x	x			dist		x
Dew and Ansari (2018)	disc			x	x			dist		
McCarthy and Fader (2018)	cont		x		x			dist		x
Xia, Chatterjee and May (2019)	cont		x		x			dist		
This paper	cont	x	x	x	x	x	x	dist	x	x

Note. *dist*, distribution for unobserved heterogeneity; *lc*, latent class for unobserved heterogeneity; *cont*, continuous setting; *disc*, discrete setting.
[a]The factor is added to the model as additional process.

factors when modeling latent customer attrition. However, none of these extensions are able to model contextual factors that vary over time.

Gupta (1991) is the first to introduce time-varying contextual factors to NBD-type models. However, in contrast to latent customer attrition models, NBD-type models focus only on customer purchases and neglect customer attrition. Recent extensions to latent customer attrition models include time-varying contextual factors for discrete and mixed noncontractual business settings have been proposed. Schweidel and Knox (2013) present a discrete probabilistic latent attrition model, which captures the purchase process with a discrete Bernoulli process and latent customer attrition with a Geometric distribution. The authors focus on time-discrete transactions in a charity setting, that is, whether a person has donated in a certain year. Gopalakrishnan et al. (2017) present a vector changepoint model in a hierarchical Bayesian framework to capture differences across a series of customer cohorts. Their approach allows for the probability of purchase and attrition to vary by individual and time.

Another stream of research focuses on adding time-varying contextual factors to the transaction process. Building on the BG/NBD model (Fader et al. 2005a), Braun et al. (2015) propose an approach for incorporating transaction-specific attributes. The authors' approach accounts for transaction characteristics at the time of purchase as contextual factors. However, these characteristics remain constant after the transaction and affect only the attrition process and not the purchase process. The authors use the model to evaluate the impact of a customer's service experience at the time of purchase. Using a MCMC approach, Harman (2016) includes time-varying contextual factors in the transaction process of the BG/NBD model. McCarthy and Fader (2018) develop an approach for customer-based corporate valuation in noncontractual settings. The authors use a beta-geometric/mixed-log-normal model and allow time-varying contextual factors to affect the purchase process.

A complementary stream of research modifies the underlying processes to address issues caused by specific contextual factors: Van Oest and Knox (2011) propose an extension of the BG/NBD model that considers customer complaints by adding a third process to the customer purchase and attrition process of the standard model. Later, Knox and van Oest (2014) add the possibility for customer recovery after complaints. Wünderlich (2015) proposes the hierarchical Bayesian seasonal model with drop-out (HSMDO), which is capable of capturing individual and cross-sectional seasonality. The foundation for this model is a discretely sampled inhomogeneous Poisson process with discrete periodic death opportunities. Platzer and Reutterer (2016) propose an extension of the Pareto/NBD model to describe regularity patterns in purchase timing. They replace the NBD distribution with a mixture of gamma distributions to allow for a varying degree of regularity across customers. The authors find empirical evidence on the importance of the regularity of timing patterns across multiple industries.

Recently, non–Pareto-type models were proposed as alternative approaches for customer base analysis. Dew and Ansari (2018) use a nonparametric framework for customer base analysis and propose the Gaussian process propensity model (GPPM). The authors use Bayesian nonparametric Gaussian priors to combine the latent functions of customer purchase behavior and further include a latent function for calendar-based time. Their model relies on a normal population distribution to control for unobserved customer heterogeneity. Xia et al. (2019) apply deep-learning algorithms in the form of conditional restricted Boltzmann machines to predict customer purchase patterns. Their approach does not distinguish between purchases and attrition but rather models the customer's purchase pattern as a whole. However, the model allows to include time-invariant and time-varying contextual factors, such as demographics and marketing activities.

3. Model

Our approach builds on the Pareto/NBD model for continuous noncontractual business settings (Schmittlein et al. 1987). To model time-varying contextual factors, we draw on Gupta (1991), who proposes an extension to NBD-type models. Such models were widely used in marketing research at this time (Jeuland et al. 1980, Schmittlein et al. 1985, Ehrenberg 1988). In contrast to latent attrition models, NBD-type models do not take customer attrition into account and focus solely on the purchase process.

In the following, we explain how to include time-varying contextual factors in the Pareto/NBD model that affect both the purchase and attrition rate. First, we introduce time-varying contextual factors into the purchase process, and second, we introduce these factors into the customer attrition process. In both cases, this will be achieved using a proportional hazard approach. Third, we add both processes to a single closed-form expression while also introducing customer heterogeneity to the two processes. Fourth, we derive the related expressions to determine the probability that a customer is alive at the end of the estimation period (in previous studies sometimes also named calibration period) (Schmittlein et al. 1987), the conditional expectation, and the discounted expected conditional transactions (DECT; analogous to the DERT proposed by Fader et al. 2005b, 2010). Finally, we discuss model identification and propose three generalized approaches to control for endogenous contextual factors. In this context, we also discuss the impact of controlling for endogenous contextual factors on predictive accuracy.

3.1. Purchase Process

Introducing time-varying contextual factors to the purchase process consists of multiple steps. First, we build on the same assumptions as the standard Pareto/NBD model, where customer purchases are modeled according to a Poisson process with exponentially distributed interpurchase times. Second, we allow the purchase rate to be a function of the time-varying contextual factors and, therefore, also time. Next, to simplify all further derivations, we follow Gupta (1991) and assume that those contextual factors are time invariant for a time interval of fixed length. Finally, we derive a general expression for the purchase process. Later, we refer to these steps to facilitate a better understanding of the model derivation. We denote variables specific for the purchase process with the superscript P.

Without time-varying contextual factors being present, the purchase behavior of a customer i is modeled as a Poisson process with rate λ_i with regard to the number of purchases made during a specified time interval (i.e., the estimation period, $(0, t]$). It follows that the probability of the number of purchases x in a time interval $(0, t]$ is given by

$$P(Y(t) = x) = \frac{[\lambda_i t]^x}{x!} \exp[-\lambda_i t]. \tag{1}$$

The probability of zero purchases for (1) defines the survivor function for the interpuchase time

$$P(Y(t) = 0) = S^P(t) = \exp(-\lambda_i t), \tag{2}$$

and therefore, the density function for the interpurchase time is given by

$$f^P(t) = \lambda_i \exp(-\lambda_i t). \tag{3}$$

This corresponds to an exponentially distributed interpurchase time for an individual customer.

To include time-varying contextual factors, we build on the proportional hazard approach. Thus, we allow the purchase rate λ_i to be a function of the contextual factors and therefore also of time:

$$\lambda_i(t) = \lambda_0 \exp\left(\gamma'_{purch} \mathbf{x}^P_{it}\right), \tag{4}$$

where $\mathbf{x}^P_{it}$ denotes the vector of contextual factors influencing the purchase process at time t, γ_{purch} represents a vector of contextual factor effect sizes, and λ_0 is the base purchase rate. For simplicity, we assume λ_0 to be identical for all customers. This assumption will be relaxed later when customer heterogeneity is introduced to the model. To simplify the notation, we remove the customer index i for all subsequent derivations. The coefficient γ_{purch} may be directly interpreted as rate elasticity. A 1% change in a contextual factor $\mathbf{x}^P$ changes the purchase rate by $\gamma'_{purch}\mathbf{x}^P$%. An increase of the purchase rate corresponds to a decrease in interpurchase time (Gupta 1991). Because $\lambda(t)$ is now a function of time-varying contextual factors, the purchase process is now a nonhomogenous Poisson process (Ross 2014). In particular, the times between the purchases of an individual customer are no longer independent and identically distributed, except if $\lambda(t) = \lambda$. It follows that the probability of the number of purchases x in a time interval $(0, t]$ is given by (1) where we substitute λt by

$$\theta^P_0(t) = \int_0^t \lambda(\tau)\, d\tau. \tag{5}$$

Note that (5) reduces to λt for a homogenous Poisson process. The survivor function and the density for the interpurchase time are given as (2) and (3) with λt substituted by (5).

In contrast to the standard Pareto/NBD model, the timing of purchases becomes relevant when introducing time-varying contextual factors. In particular, the interpurchase time between events j and $j+1$ may have a different distribution compared with the one between events $j+1$ and $j+2$. To account for this fact, we rewrite (1) using indices specifying the transactions. We denote for all subsequent derivations t_j as the time of transaction j with $j = 1, \ldots, x$ and $z_{j-1,j} = t_j - t_{j-1}$ the time passed between transaction $j-1$ and j. The probability of the number of purchases in a time interval $(t_{j-1}, t_j]$ is then given by

$$P\left(Y\left(t_{j-1}, t_j\right) = x\right) = \frac{\left[\theta^P_{t_{j-1}}\left(z_{j-1,j}\right)\right]^x}{x!} \exp\left[-\theta^P_{t_{j-1}}\left(z_{j-1,j}\right)\right], \tag{6}$$

with

$$\theta^P_{t_{j-1}}\left(z_{j-1,j}\right) = \int_{t_{j-1}}^{t_j} \lambda(\tau)\, d\tau. \tag{7}$$

With $Z_{j-1,j}$ as a random variable specifying the time between purchases $j-1$ and j, we may derive the survivor function:

$$P\left(Z_{j-1,j} > z_{j-1,j} | t_{j-1}\right) = P\left(X\left(t_{j-1}, t_j\right) = 0\right) = \exp\left[-\theta^P_{t_{j-1}}\left(z_{j-1,j}\right)\right] \tag{8}$$

and arrive at the following density function of the interpurchase time:

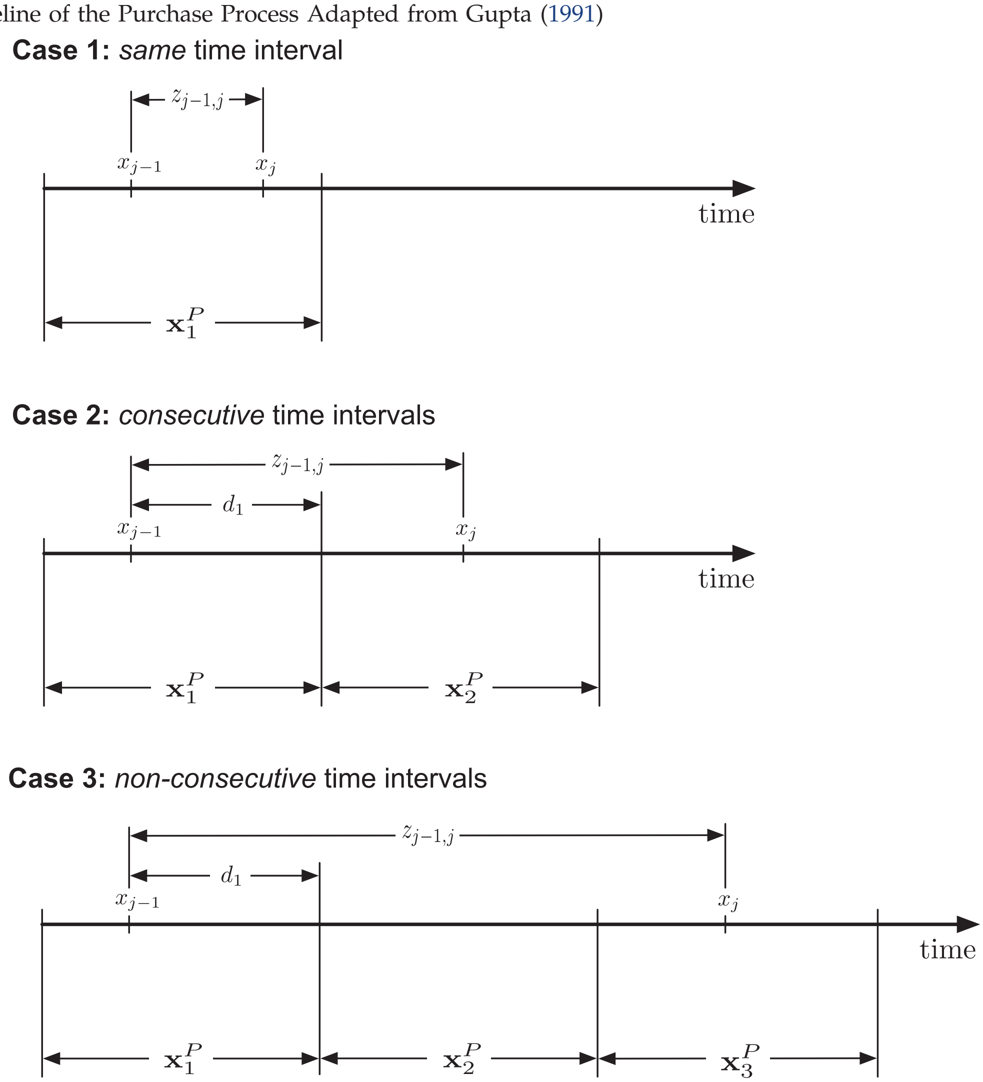
$$\begin{aligned} f^P\left(z_{j-1,j} | t_{j-1}\right) &= f^P\left(z_{j-1,j} | z_{0,1}, z_{1,2}, \ldots, z_{j-2,j-1}\right) \\ &= \lambda\left(t_j\right) \exp\left[-\theta^P_{t_{j-1}}\left(z_{j-1,j}\right)\right]. \end{aligned} \tag{9}$$

Following Gupta (1991), we simplify $\theta^P_{t_{j-1}}(t)$ by assuming that the contextual factors are time-invariant for time intervals of fixed length (e.g., one week). Therefore, the contextual factors change only for different intervals while staying constant in between. In principle, these time intervals can be arbitrarily small. It is important to emphasize that while the time-varying contextual factors are discretized, the underlying purchase process remains continuous. In consequence, $\theta^P(t)$, will consist of multiple $\lambda(t)$ components. We denote k_j as the number of intervals since purchase $j-1$. Simplifying the notation, we index the contextual factors as $x^P_1, \ldots, x^P_{k_j}$. By doing so, the factors are indexed relative to the purchases, where the interval count starts at *one* for every purchase. In particular, x_1 specifies the contextual factors in the interval that contains the $(j-1)$th transaction, whereas x_{k_j} specifies the ones in the interval of the jth transaction. d_1 is defined as the time of the transaction $(j-1)$ to the end of the first interval for any two successive transactions $(j-1)$ and j. Note that d_1 is indexed relative to the purchases. To illustrate the different components of $\theta^P_{t_{j-1}}(t)$ in (7), let us outline three possible cases. The different cases are illustrated in Figure 1.

Case 1. The two purchases that define the interpurchase time occur during the *same* time interval (e.g., the first time interval),

$$\lambda(t) = \lambda_1 = \lambda_0 \exp\left(\gamma'_{purch}\mathbf{x}^P_1\right)$$
$$\theta^P_{t_{j-1}}(t) = \int_{t_{j-1}}^{t_j} \lambda(\tau)\, d\tau = \int_0^{z_{j-1,j}} \lambda_1\, d\tau$$
$$= \lambda_1 z_{j-1,j}.$$

Figure 1. Timeline of the Purchase Process Adapted from Gupta (1991)



Case 2. The two purchases that define the interpurchase time occur in two *consecutive* time intervals (e.g., the first and the second time interval),

$$\lambda(t) = \lambda_2 = \lambda_0 \exp\left(\gamma'_{purch}\mathbf{x}_2^P\right)$$
$$\theta^P_{t_{j-1}}(t) = \int_{t_{j-1}}^{t_j} \lambda(\tau)\, d\tau = \int_{t_{j-1}}^{d_1+t_{j-1}} \lambda_1\, d\tau + \int_{d_1+t_{j-1}}^{t_j} \lambda_2\, d\tau$$
$$= \lambda_1 d_1 + \lambda_2\left(z_{j-1,j} - d_1\right).$$

Case 3. The two purchases that define the interpurchase time occur in two *nonconsecutive* time intervals (e.g., the first and the third time interval),

$$\lambda(t) = \lambda_3 = \lambda_0 \exp\left(\gamma'_{purch}\mathbf{x}_3^P\right)$$
$$\theta^P_{t_{j-1}}(t) = \int_{t_{j-1}}^{t_j} \lambda(\tau)\, d\tau = \int_{t_{j-1}}^{d_1+t_{j-1}} \lambda_1\, d\tau$$
$$+ \int_{d_1+t_{j-1}}^{1+d_1+t_{j-1}} \lambda_2\, d\tau + \int_{1+d_1+t_{j-1}}^{t_j} \lambda_3\, d\tau$$
$$= \lambda_1 d_1 + \lambda_2 + \lambda_3\left(z_{j-1,j} - d_1 - 1\right).$$

All further cases, with more time intervals between the two purchase events, are composed in the same manner. To derive a general expression for this pattern, we define

$$\lambda(t) = \lambda_0 \exp\left(\gamma'_{purch}\mathbf{x}_l^P\right), \tag{10}$$

where $l = 1, \ldots, k_j$ denotes the interval in which t lies. It follows that $\lambda(t_j) = \lambda_{k_j}$, and we arrive at a more practical version of (7):

$$\theta^P_{k_j}\left(z_{j-1,j}\right) = \lambda_1 d_1 + \sum_{l=2}^{k_j-1} \lambda_l + \lambda_{k_j}\left[z_{j-1,j} - d_1 - \delta\left(k_j - 2\right)\right], \tag{11}$$

where $\delta = 0$ if $k_j = 1$ and $\delta = 1$ if $k_j \geq 2$. δ accounts for the special case, where two subsequent purchases occur within the same time interval. Substituting $\lambda(t)$ and $\theta^P_{k_j}(z_{j-1,j})$ from (10) and (11) into (8) and (9) provides the survivor and density functions with time-varying contextual factors. Remember that contextual factors which are constant over time, result in a homogeneous process (i.e., $\lambda(t) = \lambda$ and $\theta(t) = \lambda t$).

Thus far, we have introduced time-varying contextual factors into the purchase process. In the next section, we introduce time-varying contextual factors into the attrition process.

3.2. Attrition Process

To include time-varying contextual factors in the attrition process, we proceed analogously. Although Gupta (1991) introduces contextual factors to the purchase process only, the underlying methodology is also applicable for the attrition process. In a first step, we build on the assumption of the standard Pareto/NBD model, where customer attrition is modeled by an exponential distribution. Next, we allow the attrition rate to be a function of the time-varying contextual factors and then assume that those contextual factors are time invariant for a time interval of fixed length. Finally, we derive a general expression for the attrition process. The following discussion is structured analogously to the model derivation of the purchase process. We denote variables specific for the attrition process with the superscript A.

Without time-varying contextual factors being present, the unobserved lifetime Ω of an individual customer is exponentially distributed. At the end of this lifetime, the customer is considered inactive. This leads to the following density and the survivor function for the attrition process:

$$f^A(\omega) = \mu_i \exp\left(-\mu_i \omega\right), \tag{12}$$
$$S_i^A(\omega) = P(\Omega > \omega) = \exp\left(-\mu_i \omega\right), \tag{13}$$

where μ_i is the attrition rate of an individual customer i. To compare the attrition process to the purchase process, we can express attrition as a purchase-like process with two events for every customer: the initial transaction is the first and a customer's (unobserved) attrition is the second event. Therefore, the properties and derivations of the two processes are very similar.

When including time-varying contextual factors, the attrition rate becomes a function of the contextual factors and of time:

$$\mu_i(\omega) = \mu_0 \exp\left(\gamma'_{attr}\mathbf{x}^A_{i\omega}\right), \tag{14}$$

where $\mathbf{x}^A_{i\omega}$ denotes the vector of contextual factors influencing a customer's attrition process during his lifetime ω. The variable γ_{attr} represents a vector of effect sizes for contextual factors. μ_0 is the base attrition rate. These contextual factors may or may not be the same as $\mathbf{x}^P_{it}$. Initially, we assume that μ_0 is the same for all customers. This assumption will be relaxed later. The interpretation of the effect sizes γ_{attr} is identical to the purchase process: a 1% change in a contextual factor $\mathbf{x}^A$ changes the attrition rate by $\gamma'_{attr}\mathbf{x}^A$%. Again, we remove the customer index i in all subsequent derivations to simplify the notation.

Based on the density derived for the transaction process (9), the density of a customer's lifetime Ω is given by

$$f^A(\omega) = \mu(\omega) \exp\left[-\theta^A_0(\omega)\right], \tag{15}$$

where

$$\theta^A_0(\omega) = \int_0^\omega \mu(\tau)\, d\tau. \tag{16}$$

The cumulative distribution function is given by

$$F(\omega) = \int_0^{\omega} f^A(\tau)\, d\tau = 1 - \exp[-\theta_0^A(\omega)]. \tag{17}$$

We simplify $\theta^A(\omega)$ by assuming that the contextual factors are time invariant for certain time intervals of fixed length. For simplicity reasons, we assume that the time intervals are identical to the ones from the purchase process. However, this is not required. In consequence, $\theta_0^A(\omega)$ also consists of multiple components, which are composed in the same manner as $\theta_{t_j}^P(t)$ of the purchase process. Let k_ω denote the number of the time intervals from zero to the end of a customer's lifetime ω. Then, the general case for all components of $\theta_{k_\omega}^A$ is given by

$$\theta_{k_\omega}^A(\omega) = \mu_1 d_2 + \sum_{l=2}^{k_\omega - 1} \mu_l + \mu_{k_\omega}[\omega - d_2 - \delta(k_\omega - 2)], \tag{18}$$

where $\mu_l = \mu_0 \exp(\gamma'_{attr}\mathbf{x}_l^A)$ and $\delta = 0$ if $k_\omega = 1$ and $\delta = 1$ if $k_\omega \geq 2$, d_2 is defined as the time between 0 and the end of the first interval. Finally, δ accounts for the special case where a customer's first purchase and attrition occur in the same time interval.

3.3. Likelihood Derivation

Thus far, we modeled the customer purchase and attrition processes separately. To derive the individual likelihood, we combine the two processes. This consists of multiple steps. First, we take a closer look at the purchase process in order to identify different ways customer attrition can affect it. Then, we combine the purchase and the attrition process. Next, we introduce customer heterogeneity and relax the assumption of identical purchase and attrition rates across customers. Finally, we arrive at a closed-form solution for the individual likelihood. Our approach uses a similar methodology as the one illustrated by Fader and Hardie (2005) for the standard Pareto/NBD model. In this section, we discuss the steps for deriving the individual likelihood; more detailed mathematical derivations are presented in Appendix A.1 and Appendix B.

Before combining the purchase and the attrition process, we need to distinguish two cases that affect the purchase process. When observing a customer during the estimation period $(0, T]$, the following scenarios are possible: either (1) he is alive at the end of the estimation period T or (2) the customer becomes inactive between the last observed transaction and T. We compose the individual likelihood as the product of the densities for the interpurchase times and the survivor function. With regard to the latter, we use the survivor function between the first transaction after the end of the estimation period T and the very last transaction of the customer (note that both of these transactions are unobserved).

Case 1. The customer is still alive at the end of the estimation period T. This means that the lifetime of the customer is longer than the observed period (i.e., $\omega > T$). To compose the individual likelihood in this case, the survivor function is evaluated at the time between the last observed transaction and the end of the estimation period.

Case 2. The customer becomes inactive between the last transaction and T. This means that the customer's lifetime ends somewhere before the end of the estimation period (i.e., $\omega \in (t_x, T]$). To compose the individual likelihood in this case, the survivor function is evaluated at the time between the last observed transaction and the end of the customer's lifetime ω.

In a next step, we combine the two cases and derive the individual likelihood conditional on λ_0 and μ_0:

$$\begin{aligned} &L(\lambda_0, \gamma_{purch}, \gamma_{attr}, \mu_0 | \mathbf{t}, T, \mathbf{X}^P, \mathbf{X}^A, x) \\ &\quad = L(\lambda_0, \gamma_{purch} | \mathbf{t}, T, \mathbf{X}^P, x, \Omega > T) P(\Omega > T | \mu_0) \\ &\qquad + \int_{t_x}^{T} L(\lambda_0, \gamma_{purch} | \mathbf{t}, T, \mathbf{X}^P, x, \omega \in (t_x, T]) \\ &\qquad \times f^A(\omega | \mu_0)\, d\omega. \end{aligned} \tag{19}$$

The integral in expression (19) depends on ω through $k_{x,\omega}$. To solve it, we use the assumption that the contextual factors are constant during fixed time intervals. Therefore, we are able to split the integral into parts, where $k_{x,\omega}$ is constant, with $k_{x,\omega} = 1, \ldots, k_T$.

Finally, we relax the assumption that all customers have the same purchase and attrition rate and introduce customer heterogeneity (i.e., we allow λ_0 and μ_0 to vary across customers). Following Schmittlein et al. (1987), we assume the purchase rate Λ_0 and the attrition rate M_0 to be Gamma distributed with shape parameter r and scale parameter α, respectively, s and β,

$$g(\lambda_0) = \frac{\alpha^r \lambda_0^{r-1} e^{-\lambda_0 \alpha}}{\Gamma(r)} \text{ and } g(\mu_0) = \frac{\beta^s \mu_0^{s-1} e^{-\mu_0 \beta}}{\Gamma(s)}$$

and arrive at

$$\begin{aligned} &L(\alpha, r, \beta, s, \gamma_{purch}, \gamma_{attr} | \mathbf{t}, T, \mathbf{X}^P, \mathbf{X}^A, x) \\ &\quad = \int_0^{\infty} \int_0^{\infty} L(\lambda_0, \mu_0, \gamma_{purch}, \gamma_{attr} | \mathbf{t}, T, \mathbf{X}^P, \mathbf{X}^A, x) \\ &\qquad \times g(\lambda_0) g(\mu_0) d\lambda_0 d\mu_0. \end{aligned} \tag{20}$$

By removing the conditioning on λ_0 and μ_0, we are able to derive the final individual likelihood. For the closed-form solution of (20), see Appendix A.1. The detailed steps for deriving the closed-form solution including solving the integral in (20) are provided in Appendix B.

3.4. P(alive)

In noncontractual settings, customer churn is not observed. Therefore, the probability of being alive at the end of the estimation period T, P(alive), is a key component of predicting future purchase behavior. P(alive) is the probability that the lifetime of the customer is larger than T, $P(\Omega > T)$, given all available information. To obtain P(alive), we first apply the Bayes' theorem to obtain the posterior probability of being alive. Because the purchase rate λ_0 and the attrition rate μ_0 are not directly observed, we further need to include the posterior distribution of (λ_0, μ_0), before we can integrate them out and arrive at the closed-form solution for P(alive).

Applying Bayes' theorem to the individual likelihood (19) results in

$$P(\Omega > T|\lambda_0, \mu_0, \mathbf{X}^P, \mathbf{X}^A, \gamma_{purch}, \gamma_{attr}, x, \mathbf{t}, T) = \frac{L(\lambda_0, \gamma_{purch}|\mathbf{X}^P, x, \mathbf{t}, T, \Omega > T)P(\Omega > T|\mu_0)}{L(\lambda_0, \gamma_{purch}, \gamma_{attr}, \mu_0|\mathbf{t}, T, \mathbf{X}^P, \mathbf{X}^A, x)}. \quad (21)$$

Because the purchase rate λ_0 and the attrition rate μ_0 are not directly observed, we multiply (21) by the posterior distribution of (λ_0, μ_0), which is given as

$$g(\lambda_0, \mu_0|r, \alpha, s, \beta, \gamma_{purch}, \gamma_{attr}, \mathbf{X}^P, \mathbf{X}^A, x, \mathbf{t}, T) = \frac{L(\lambda_0, \gamma_{purch}, \gamma_{attr}, \mu_0|\mathbf{X}^P, \mathbf{X}^A, x, \mathbf{t}, T)g(\lambda_0)g(\mu_0)}{L(\alpha, r, \beta, s, \gamma_{purch}, \gamma_{attr}|\mathbf{X}^P, \mathbf{X}^A, x, \mathbf{t}, T)}. \quad (22)$$

Then, we take the integral with respect to this joint distribution to obtain the final P(alive) expression:

$$\begin{aligned} &P(\Omega > T|r, \alpha, s, \beta, \gamma_{purch}, \gamma_{attr}, \mathbf{X}^P, \mathbf{X}^A, x, \mathbf{t}, T) \\ &= \int_0^\infty \int_0^\infty P(\Omega > T|\lambda_0, \mu_0, \mathbf{X}^P, \mathbf{X}^A, \gamma_{purch}, \gamma_{attr}, x, \mathbf{t}, T) \\ &\quad \times g(\lambda_0, \mu_0|r, \alpha, s, \beta, \gamma_{purch}, \gamma_{attr}, \mathbf{X}^P, \mathbf{X}^A, x, \mathbf{t}, T) \\ &\quad \times d\lambda_0 d\mu_0. \end{aligned} \quad (23)$$

The closed-form solution to (23) is reported in Appendix A.2, and related derivations are reported in Appendix C.

3.5. Conditional Expectation

The conditional expectation metric allows practitioners and researchers to predict the future number of purchases of an individual given a customer's past purchase behavior. To derive the expression, we follow multiple steps. We start with the assumption that the customer did not stop purchasing before the prediction period. However, we still have to distinguish two cases for a customer's possible attrition during the prediction period. We use the characteristics of nonhomogenous Poisson processes and the fact that the contextual factors are assumed to be time invariant for an interval of fixed length to derive the first conditional expectation. Last, we relax our initial assumption and allow attrition before the prediction period. In this section, we discuss the steps for deriving the conditional expectation; more mathematical derivations are presented in Appendix C.

We start by deriving $E[Y(T, T+t)|r, \alpha, s, \beta, \gamma_{purch}, \gamma_{attr}, \mathbf{X}^P, \mathbf{X}^A, x, \mathbf{t}, T]$, where $Y(T, T+t)$ is a random variable specifying the number of purchases made after the estimation period in the interval $(T, T+t]$. Our approach is based on Fader and Hardie (2005). We start deriving the conditional expectation under the assumption that the customer is alive after the end of the estimation period T. This assumption is relaxed later. Under this assumption, the conditional expectation is a special case of the purchase process, where we have two transactions, the first one at T and the second one at ω, with $\omega \in (T, T+t]$.

Although we assume that all customers are alive at the end of the estimation period, there are still two different cases to consider about the timing of customer attrition. The customer may stop buying after the end of the prediction period or during the prediction period.

Case 1. The lifetime of the customer is longer than the prediction period, that is, $\omega > T + t$. In this case, we get the expectation of the nonhomogenous purchase process with $\omega > T + t$:

$$E[Y(T, T+t)|\lambda_0, \mathbf{X}^P, \gamma_{purch}, \Omega > T+t].$$

Case 2. The customer stops purchasing between the end of the observation period and the end of the prediction period. In this case, we get the expectation of the nonhomogenous attrition process with $\omega \in (T, T+t]$:

$$E[Y(T, T+t)|\lambda_0, \mathbf{X}^P, \gamma_{purch}, \Omega \in (T, T+t]].$$

We combine the two cases, but because we do not observe the customer lifetime ω, we also have to consider the attrition process:

$$\begin{aligned} &E[Y(T, T+t)|\lambda_0, \mu_0, \mathbf{X}^P, \mathbf{X}^A, \gamma_{purch}, \gamma_{attr}, \Omega > T] \\ &= E[Y(T, T+t)|\lambda_0, \mathbf{X}^P, \gamma_{purch}, \Omega > T+t] \\ &\quad \times P(\Omega > T+t|\mu_0, \Omega > T) + \\ &\quad + \int_T^{T+t} \left(E[Y(T, T+t)|\lambda_0, \mathbf{X}^P, \gamma_{purch}, \omega \in (T, T+t]]\right) \\ &\quad \times f^A(\omega|\mu_0, \Omega > T)\, d\omega. \end{aligned} \quad (24)$$

For expression (24), we assume that customers are alive at the end of the observation period T, that is, $\omega > T$.

To relax this assumption, we add the posterior distribution of being alive at T, P(alive), from Equation (A.2).

$$\begin{aligned} &E[Y(T,T+t)|\lambda_0,\mu_0,\mathbf{X}^P,\mathbf{X}^A,\gamma_{purch},\gamma_{attr},x,\mathbf{t},T] \\ &\quad = E[Y(T,T+t)|\lambda_0,\mu_0,\mathbf{X}^P,\mathbf{X}^A,\gamma_{purch},\gamma_{attr},x,\mathbf{t},\Omega > T] \\ &\quad\quad \times P(\Omega > T|\lambda_0,\mu_0,\mathbf{X}^P,\mathbf{X}^A,\gamma_{purch},\gamma_{attr},x,\mathbf{t},T). \qquad (25) \end{aligned}$$

By removing the conditioning on λ_0 and μ_0, we arrive at the closed-form solution reported in Appendix A.3.

3.6. DECT

To estimate individual customer lifetime values with the proposed model in combination with a Gamma/Gamma model (Colombo and Jiang 1999, Fader et al. 2005b), we need to derive the present value of the expected future transactions and account for the discount rate. We define the DECT. The concept of this metric is very similar to the discounted expected residual transactions (DERT) (Fader et al. 2005b, 2010).

The general formulation of the customer lifetime value (CLV) is

$$E(CLV) = \int_0^\infty v(t)S(t)d(t)dt,$$

where $v(t)$ is the customer value at t, $S(t) = 1 - F(t) = P(\Omega > t)$ denotes the survivor function of Ω and $d(t)$ a specific discount factor. Assuming that the value of each transaction remains constant, $v(t)$ can be factored out leaving only the purchase rate $\lambda(t)$ in the integral:

$$\int_0^\infty \lambda(t)S(t)d(t)dt.$$

These are the discounted expected transactions (DETs).

Starting instead at $t = T$, the end of the estimation period, for continuous discounting with a rate of Δ, gives the DERT (Fader et al. 2005b, 2010):

$$\begin{aligned} &\mathrm{DERT}(\Delta|\lambda_0,\mu_0,\mathbf{X}^P,\mathbf{X}^A,\gamma_{purch},\gamma_{attr},\Omega > T) \\ &\quad = \int_T^\infty \lambda(t)P(\Omega > t|\Omega > T)d(t)\,dt \\ &\quad = \int_T^\infty \lambda_0 \exp(\mathbf{x}_t^P)\exp[-\mu_0(\theta_0(t) - \theta_0(T))] \\ &\quad\quad \times \exp[-\Delta(t-T)]\,dt. \end{aligned}$$

When modeling time-varying contextual factors, we face the issue that the measurements of the time-varying contextual factors are likely not known up to infinity. Thus, we make predictions only within a time frame for which the time-varying contextual factors are known. Therefore, we truncate the time horizon of our predictions and approximate the DERT as DECT:

$$\begin{aligned} &\mathrm{DECT}(\Delta,t|r,\alpha,s,\beta,\gamma_{purch},\gamma_{attr},\mathbf{X}^P,\mathbf{X}^A,x,\mathbf{t},T) = \\ &\quad = \int_T^{T+t} \lambda_0 \exp(\mathbf{x}_\tau^P)\exp[-\mu_0(\theta_0(\tau) - \theta_0(T))] \\ &\quad\quad \times \exp[-\Delta(\tau - T)]\,d\tau, \qquad (26) \end{aligned}$$

where $t > 0$ is the prediction length. We present the closed-form solution to the integral in (26) in Appendix A.4.

3.7. Model Identification

When considering contextual factors in the Pareto/NBD model, model identification has to be addressed. In the following, we focus on the ability of the extended Pareto/NBD model to separately identify the effects of time-varying contextual factors on the purchase process and the attrition process. The identification of standard latent customer attrition models without contextual factors has been thoroughly discussed in prior studies (Schmittlein et al. 1987, Fader and Hardie 2005, Schweidel and Knox 2013).

First, we describe the effects of contextual factors: modeling contextual effects means that a part of the unobserved heterogeneity is substituted by systematic differences, which are expressed by the contextual factors. As such, the parameter r, which measures homogeneity in the purchase process, and the parameter s, which measures homogeneity in the attrition process, should increase when modeling contextual effects (Gupta 1991). The effects of contextual factors may influence both the purchase and the attrition processes, just one of the processes, or neither of the processes. Although this applies to both time-invariant and time-varying contextual factors, the identification of these two categories of contextual effects relies on different sources of variation.

To identify time-invariant contextual factors, it is possible to leverage the variation across customer groups. For example, female customers might purchase more. However, their loyalty is the same as that of male customers. Thus, gender will influence only the purchase process and not the attrition process. Another example is that older customers might be more loyal to the retailer, and, therefore, more time must pass until they stop purchasing. However, their purchase rate may be the same as that of younger customers.

The identification of time-varying contextual factors, which allow us to capture the effects of dynamic changes in a customer's context (e.g., personalized marketing), is more complex. Individual customers are allowed to change their purchase and attrition behavior over time. Temporal events may cause a strictly temporary change in the purchase rate (e.g., shortly after receiving a personalized marketing offer, customers tend to purchase more). However, such an event may also influence a customer's future purchase behavior. For example, personalized marketing might encourage customers to return to the store in the future and therefore can reduce the likelihood that a customer will stop purchasing at this point in time. To identify such patterns, variation across time must be considered.

Building on a comprehensive simulation study, we analyze various scenarios and find support for the identification of contextual factors in the extended Pareto/NBD model. The fitted model recovers the true parameters for the contextual factors of the simulated data. The simulation study was conducted as follows: we generate transactional data of 3,000 customers for Poisson-distributed purchases and incorporate the exponentially distributed attrition assumption including effects for time-varying contextual factors. Thereby, a contextual factor may affect the two underlying processes in four different ways: (1) only the purchase process is affected; (2) only the attrition process is affected; (3) both processes are affected; or (4) the contextual factor does not have an effect on either process. We generate contextual factors as both a binary variable (i.e., dummy variables for seasonal patterns) and a count variable (i.e., direct marketing actions) ranging from $[0, 16]$. The true effect sizes of the contextual factors vary across three levels (0.5, 1, and 1.5). These effect sizes may be interpreted as rate elasticities for the purchase and attrition rates. To test the model identification for different levels of homogeneity across customers for both the purchase rate and attrition rate, we vary the shape parameters (r and s, respectively) between a high and a low level of customer homogeneity (high $r = 4.5, s = 1.5$ and low $r = 1.5, s = 0.5$). The scale parameters (α and β) are kept constant ($\alpha = 170, \beta = 8$). Both the shape and scale parameters are set to correspond to the values observed in the empirical analyses of real-world data sets in this paper (see Section 4). In total, we analyze 48 scenarios. Every scenario is simulated 50 times to account for sample variation in the data generation process.

To determine the effect sizes, we apply the extended Pareto/NBD model to the simulated datasets over a two-year estimation period. Table 2 shows the results of the simulation study. The reported figures are mean values of 50 repeated simulations. The model is capable of accurately recovering the parameter values for all settings. Notably, we observe a greater variance for the attrition process than for the purchase process. This occurs because the customer attrition rate is, in contrast to the purchase rate, not directly observed in the model. In addition, we also observe a greater variance for the binary variable than for the count variable. This is expected, because binary factors provide less information.

3.8. Endogenous Contextual Factors and Recovering the True Parameter Values

The extended Pareto/NBD model assumes that the contextual factors are exogenous. Although this is true for contextual factors such as seasonal patterns, marketing activities are nonrandom in most cases, as they are often based on an individual customer's previous performance (Shugan 2004, Schweidel and Knox 2013). To accurately quantify effects of marketing activities and similar contextual factors, we have to account for their nonrandom nature. In the following, we discuss three different approaches to control for endogenous contextual factors in the extended Pareto/NBD model for the purchase process. The approaches are identical for the attrition process. We start discussing an IV approach, followed by a control function and finally a copula correction method. Furthermore, we present the results of an additional simulation study that provides evidence of the capability of the extended Pareto/NBD model to recover the actual effect sizes of endogenous contextual factors.

Let us consider the purchase process of one randomly chosen customer with a mean purchase rate given as

$$\Lambda_k = \Lambda_0 \exp\left(\gamma_{purch} \mathbf{x}_k\right), \tag{27}$$

where $k = 1, \ldots, K$ denotes the predefined time intervals for the contextual factors. For simplicity, we assume there is only one contextual factor. In interval k, the number of transactions is distributed according to a standard homogenous Poisson process. We can rewrite (27) as

$$\log(\Lambda_k) = \gamma_{purch} \mathbf{x}_k + \log(\Lambda_0). \tag{28}$$

The heterogeneity in Λ_0 is Gamma distributed with the parameters r and α; therefore, the density of $N := \log(\Lambda_0)$ can be derived as

$$f(\nu) = \frac{\alpha^r}{\Gamma(r)} \exp\left(r\nu - \alpha \exp(\nu)\right), \tag{29}$$

which is known as the Loggamma distribution: $N \sim \text{LogGam}(r, \alpha)$. Accordingly, $\log(\Lambda_k)$ is log-gamma distributed with an additional location term $\gamma_{purch} \mathbf{x}_k$:

$$\log(\Lambda_k) = \gamma_{purch} \mathbf{x}_k + N_k.$$

If the contextual factor is endogenous, $\mathbf{x}_k$ is correlated with N_k. If there is a positive correlation, a high base transaction rate λ_0 will lead to a high $v = \log(\lambda_0)$ and consequently, to a high $\mathbf{x}_k$ biasing γ_{purch}.

To address this issue, we introduce an IV $\mathbf{z}_k$ that correlates with $\mathbf{x}_k$ but is independent of N_k:

$$\mathbf{x}_k = \beta_0 + \beta_1 \mathbf{z}_k + \epsilon_k, \tag{30}$$

where ϵ_k is independent and identically distributed and correlated with the Loggamma distribution of N_k. In the case of marketing activities, previous research (Schweidel and Knox 2013) relies on lagged recency, frequency, and monetary value (RFM) scores to model

Table 2. Simulation Results for Recovering Exogenous Contextual Factors

	Contextual factor type	Homogeneity	True value γ_{purch}	Mean estimate γ_{purch}	Estimate SD γ_{purch}	True value γ_{attr}	Mean estimate γ_{attr}	Estimate SD γ_{attr}
Contextual factor influences both processe	Binary	High	0.5	0.516	0.081	0.5	0.555	0.135
	Binary	High	1	0.989	0.072	1	0.990	0.149
	Binary	High	1.5	1.473	0.074	1.5	1.538	0.118
	Binary	High	0.5	0.482	0.106	1.5	1.539	0.149
	Binary	High	1.5	1.488	0.065	0.5	0.518	0.060
	Binary	Low	0.5	0.488	0.070	0.5	0.506	0.241
	Binary	Low	1	0.998	0.066	1	0.968	0.285
	Binary	Low	1.5	1.509	0.073	1.5	1.439	0.312
	Binary	Low	0.5	0.516	0.085	1.5	1.462	0.313
	Binary	Low	1.5	1.516	0.070	0.5	0.553	0.209
	Count	High	0.5	0.516	0.045	0.5	0.527	0.071
	Count	High	1	1.006	0.043	1	1.043	0.071
	Count	High	1.5	1.436	0.019	1.5	1.483	0.089
	Count	High	0.5	0.480	0.095	1.5	1.452	0.114
	Count	High	1.5	1.433	0.029	0.5	0.506	0.058
	Count	Low	0.5	0.480	0.095	0.5	1.452	0.114
	Count	Low	1	0.999	0.033	1	0.975	0.120
	Count	Low	1.5	1.454	0.028	1.5	1.429	0.092
	Count	Low	0.5	0.496	0.051	1.5	1.450	0.157
	Count	Low	1.5	1.454	0.019	0.5	0.492	0.087
Contextual factor influences only purchase process	Binary	High	0.5	0.488	0.065	0	0.002	0.125
	Binary	High	1	1.018	0.066	0	−0.002	0.108
	Binary	High	1.5	1.508	0.063	0	0.032	0.093
	Binary	Low	0.5	0.512	0.072	0	0.056	0.179
	Binary	Low	1	0.999	0.061	0	0.011	0.227
	Binary	Low	1.5	1.500	0.051	0	−0.017	0.186
	Count	High	0.5	0.507	0.040	0	0.020	0.083
	Count	High	1	0.995	0.028	0	0.023	0.067
	Count	High	1.5	1.414	0.018	0	0.016	0.057
	Count	Low	0.5	0.505	0.039	0	0.004	0.163
	Count	Low	1	0.964	0.023	0	0.005	0.094
	Count	Low	1.5	1.449	0.019	0	−0.024	0.095
Contextual factor influences only attrition process	Binary	High	0	−0.008	0.111	0.5	0.528	0.172
	Binary	High	0	0.021	0.102	1	1.012	0.174
	Binary	High	0	−0.039	0.121	1.5	1.467	0.179
	Binary	Low	0	0.007	0.095	0.5	0.054	0.261
	Binary	Low	0	0.002	0.085	1	0.925	0.271
	Binary	Low	0	−0.006	0.098	1.5	1.451	0.345
	Count	High	0	0.006	0.066	0.5	0.507	0.077
	Count	High	0	−0.002	0.095	1	0.984	0.117
	Count	High	0	−0.046	0.136	1.5	1.486	0.129
	Count	Low	0	−0.006	0.052	0.5	0.472	0.193
	Count	Low	0	0.014	0.061	1	1.013	0.178
	Count	Low	0	0.005	0.067	1.5	1.421	0.212
No influence	Count	High	0	0.001	0.059	0	0.002	0.102
	Count	Low	0	0.010	0.050	0	0.058	0.161
	Binary	High	0	−0.007	0.080	0	−0.009	0.100
	Binary	Low	0	0.018	0.091	0	0.068	0.268

a firm's targeting decision. This leads us to the following two-step approach: first, we estimate (30) using least squares to obtain $\hat{\mathbf{x}}_k := \hat{\beta}_0 + \hat{\beta}_1 \mathbf{z}_k$. Second, we estimate the extended Pareto/NBD model, but instead of using the endogenous contextual factor $\mathbf{x}_k$, we use $\hat{\mathbf{x}}_k$. The procedure is similar for the attrition process. The proposed two-step IV approach assumes a linear combination of the explanatory variables (i.e., linear-in-parameters model). A second method to control for endogenous variables is the control function approach (Rivers and Vuong 1988, Petrin and Train 2010). It offers a parsimonious way to account for endogeneity of a contextual factor even if it interacts with other exogenous variables (Wooldridge 2015). Instead of estimating the extended Pareto/NBD model with $\hat{\mathbf{x}}_k$, we add the error term ϵ_k in (30) as an additional contextual factor, to the model (in addition to the endogenous contextual factor).

A third approach that can be used to control for endogenous variables is based on the copula correction method proposed by Park and Gupta (2012). Similar to control function approach, this approach adds a factor to the model. However, this additional factor is derived using a Gaussian copula and the marginal distribution of the endogenous factor. This means, in addition to the endogenous contextual factor $\mathbf{x}_k$, we add the contextual factor $\mathbf{p}_k^*$, where

$$\mathbf{p}_k^* = \Phi^{-1}(H(\mathbf{x}_k)), \tag{31}$$

with $H(\mathbf{x}_k)$ as the marginal distribution of the endogenous regressor and Φ^{-1} as the inverse cumulative distribution function of a standard normal distribution. Including the marginal distribution of the endogenous contextual factors solves the correlation between the endogenous contextual factor and the structural error and therefore results in consistent parameter estimates. Although it is more complex, the copula correction method has two key advantages: (1) there is no need to add an instrumental variable as all required information is extracted from the already available data and (2) it is applicable to discrete endogenous variables. A drawback of the copula approach is the possibility of introducing multicollinearity to the model, which may affect model convergence. To ensure identification, the discrete endogenous contextual factor must not be binomially distributed (Park and Gupta 2012).

The simulation study indicates that all three proposed approaches can recover the true parameter values of endogenous contextual factors. The simulation study was conducted as follows: we generate transactional data of 3,000 customers following the Poisson-distributed purchase and exponentially distributed attrition assumptions. In addition, using a Gaussian Copula approach, we simulate a contextual factor that is correlated with the mean purchase rate, the mean attrition rate, or both. The contextual factor is continuous for the IV and the control function approach and discrete for the copula approach. The base model parameters are kept constant ($r = 1.5, \alpha = 170, s = 0.5, \beta = 8$), corresponding to the values observed in the empirical analyses in this paper (see Section 4). We then fit the extended Pareto/NBD model over a two-year estimation period. By naively modeling this contextual factor or wrongly assuming its exogeneity, the model is not capable of recovering the true γ_{purch} or γ_{attr} (as indicated by the standard estimate column in Table 3). Next, we use the instrumental variable, control function, and copula approach proposed above to recover the actual values for γ_{purch} or γ_{attr}. Table 3 shows the results for the parameter estimates for the proposed methods. All three approaches are capable of recovering the true parameter values. As expected, the standard deviation of the attrition process is larger compared with the purchase process.

3.9. Endogenous Contextual Factors and Predictive Accuracy

Controlling for endogeneity is essential for correctly measuring and understanding the effects of the contextual factors on the two underlying processes. However, in the simulation study, we observe that controlling for endogeneity results in less accurate predictive performance for the holdout sample. Given a scenario with an endogenous contextual factor, we derive the conditional expectation metric for a two-year period. Table 4 shows the mean absolute error (MAE) and its standard deviation (SD) for individual predictions in three settings following different modeling approaches for the estimation and prediction sample. In all three cases, the contextual factor is endogenous ($corr = 0.8$). The first setting uses the two-step IV

Table 3. Simulation Results for Recovering Endogenous Contextual Tactors

	Endo. biased process	True value γ_{purch}	IV estimate γ_{purch}	IV estimate SD γ_{purch}	Standard estimate γ_{purch}	Standard estimate SD γ_{purch}	True value γ_{attr}	IV estimate γ_{attr}	IV estimate SD γ_{attr}	Standard estimate γ_{attr}	Standard estimate SD γ_{attr}
IV	Both	0.5	0.502	0.052	0.700[a]	0.044	0.5	0.469	0.124	1.037[a]	0.162
	Purch	0.5	0.495	0.062	0.687[a]	0.041	0.5	—	—	0.457	0.106
	Attr	0.5	—	—	0.495	0.062	0.5	0.458	0.106	1.128[a]	0.238
CF	Both	0.5	0.491	0.025	0.700[a]	0.044	0.5	0.453	0.310	1.037[a]	0.162
	Purch	0.5	0.494	0.030	0.687[a]	0.041	0.5	—	—	0.454	0.090
	Attr	0.5	—	—	0.495	0.025	0.5	0.501	0.345	1.128[a]	0.238
Copula	Both	0.5	0.483	0.053	0.700[a]	0.044	0.5	0.494	0.201	1.037[a]	0.162
	Purch	0.5	0.507	0.045	0.687[a]	0.041	0.5	—	—	0.459	0.041
	Attr	0.5	—	—	0.498	0.040	0.5	0.566	0.314	1.128[a]	0.238

Note. SD, standard deviation.
[a]Subject to endogeneity.

Table 4. Simulation Results for Predictive Accuracy with Endogenous Contextual Factors

Two-year prediction period	MAE	Standard deviation of MAE
Approach used IV only during estimation	0.347	0.107
Approach used IV during estimation and prediction	0.130	0.020
Approach did not use IV	0.125	0.010

approach to correct for endogenous contextual factors for the model estimation, but no IV is used to predict customer behavior during the holdout period. In the second setting, we control for endogenous contextual factors during both model estimation and prediction. In the third setting, we naively use the endogenous factor to estimate the model and predict customer behavior, that is, we do not control for endogeneity. The prediction period is two years. We observe that the setting that does not control for endogenous contextual factors outperforms the other two approaches. These findings are in line with those in Ebbes et al. (2011). Consistent with these authors, we suggest that researchers carefully think about their key model objective and thus decide whether controlling for endogenous contextual factors is advisable in their specific research context. Although it is rarely advisable if the study's focus is primarily predictive modeling, controlling for endogeneity is of utmost importance if the study's focus is on disentangling causal relationships.

4. Empirical Analysis

We use three real-world data sets from the retailing industry to test the extended Pareto/NBD model. The first data set is from a multichannel catalog merchant, the second data set is from an electronic retailer, and the third data set is from a sporting goods retailer. The first two data sets are publicly available.

Although all three datasets are from the retailing industry, they have distinct characteristics that relate to commonly observed scenarios across various industry groups. First, the nature of available time-varying contextual factors differs across firms. Some firms can rely on detailed information about contextual factors for individual customers. This is not the case for other firms, which either do not store such records or are not allowed to use them for customer base analyses, for example, because of data privacy regulations. Second, customer transactions across firms are characterized by varying interpurchase times. Although some firms record multiple transactions for individual customers each month, the time between the transactions of individual customers is considerably longer for other firms. To account for these varying characteristics, we analyze data sets that include three commonly observed scenarios in this context. The first data set corresponds to the scenario where firms have access to a wide set of contextual factors for each individual and at the aggregate level. Furthermore, the transactions for customers occur rather frequently. The second data set is widely comparable to the first data set, but information on contextual factors is limited. This means that information is only available for aggregate-level time varying contextual factors. Additionally, interpurchase times are longer than those in the first data set. The third data set again includes information on both individual and aggregate-level contextual factors and is characterized by longer interpurchase times. Analyzing multiple datasets facilitates a more robust discussion of when and how modeling contextual factors contribute to increasing the predictive accuracy of latent probabilistic customer attrition models.

First, we compare our extended Pareto/NBD model, which includes time-varying and time-invariant contextual factors, with the standard Pareto/NBD model, which does not account for any contextual factors. We compare the parameter estimates and the in-sample performance of these two models. To evaluate the in-sample performance during the estimation period, we use the Bayesian information criteria (BIC) and the log-likelihood (LL) values. Because these two models are nested, the parameters and LL values are directly comparable. Second, for the customer base analysis, we benchmark the out-of-sample predictive performance at the individual and aggregate level and compare it with that of the standard Pareto/NBD model and four state-of-the-art models. We compare it with that of the pareto/GGG model (Platzer and Reutterer 2016), which is the latest published Pareto-type model, and the transaction attribute model (TAM), which was introduced by Braun et al. (2015). Moreover, we apply the GPPM, a Bayesian nonparametric framework proposed by Dew and Ansari (2018), which is the latest published non–Pareto-type approach used for customer base analysis. Finally, we benchmark it against the latent attrition model with direct marketing activities (LAMDMA) proposed by Schweidel and Knox (2013). The parameter estimates are obtained either by maximum likelihood estimation or by using a MCMC-based approach in accordance with the original studies. Comparisons are made for each of the three datasets. To evaluate out-of-sample performance at the individual level, we use the mean absolute error (MAE) and the correlation between the predicted and observed number of transactions for the individual customers. The MAE is based on the absolute difference between the predicted number of transactions and the observed transactions up to the specified time point for every customer. We report the mean of these errors across individual customers. Similarly, the individual-level correlation measures the relationship between the predicted number of transactions and the observed number of transactions of individual customers up to the specified time point. We use the conditional expectation metric. To examine the robustness of our findings, we assess the predictive accuracy of the models for different prediction periods. Furthermore, we compare model performance in terms of identifying the future best 10% and 20% of customers (Wübben and Wangenheim 2008). We measure MAE at the aggregate level for both the predicted and observed weekly cumulative transactions.

4.1. Multichannel Catalog Merchant

The first data set is from a multichannel catalog merchant. We analyze the purchase history of 1,402 customers who made a purchase for the first time between January 2005 and March 2005. The data set is publicly available from Marketing EDGE (Marketing EDGE 2013). The customer cohort accounts for 2,929 transactions between January 2005 and September 2012. A transaction record consists of the purchase date and customer ID. Table 5 summarizes the key descriptive statistics. We fit the models based on repeated transactions during a one-year estimation period and examine the predictive performance of all four models based on different prediction periods.

In the Extended Pareto/NBD model, we consider three contextual factors: (1) catalog mailings, (2) seasonal purchase patterns, and (3) acquisition channels. Although the first two contextual factors are time varying on a weekly basis, the last is time invariant, that is, does not change over time. Customers receive multiple catalogs by mail every year, with timing of these mailings differing across customers. Catalog mailings are sent to the customers irrespective of their previous purchase histories. We include catalog mailings as an individual-level time-varying contextual factor in the extended Pareto/NBD model. The variable is operationalized as a dummy variable. Furthermore, like many retailing companies, this multichannel merchant is subject to seasonal purchase patterns. During the Christmas season, the firm observes a significant increase in purchases, for example. We include the seasonal pattern as a time-varying contextual factor at the aggregate level, that is, although varying over time, this contextual factor is the same for all customers. The information on high season is derived from historical data and expert knowledge. Finally, we distinguish between online and offline customers and include the first-purchase channel for every customer as a time-invariant contextual factor.

Our results show that the extended Pareto/NBD model has a better in-sample fit in terms of the LL and BIC. The parameter estimates and model fit are shown in Table 6. The frequency plot of repeat transactions in Figure 2 indicates a good in-sample fit. Because heterogeneity in the data are captured by Gamma distributions, the parameters r and s are direct measures of the extent of heterogeneity in customers' purchase and attrition rates, respectively. The higher the value of parameter r, the higher the homogeneity in the purchase process, and the higher the value of parameter s, the higher the homogeneity in the attrition process (Gupta 1991). Explicitly modeling contextual factors in the extended Pareto/NBD model explains unobserved heterogeneity among customers. In the extended Pareto/NBD, we observe higher values for r and s, indicating more homogeneity among customers for both processes.

The coefficients of the contextual factors may be directly interpreted as rate elasticity. A 1% change in a contextual factor $\mathbf{X}^P$ or $\mathbf{X}^A$ changes the purchase or attrition rates by $\gamma_{purch}\mathbf{X}^P$ or $\gamma_{attr}\mathbf{X}^A$%, respectively (Gupta 1991). Every contextual factor has two coefficients. While the first coefficient represents the influence on the transaction rate, the second coefficient represents the effects on the customer's lifetime. The coefficients of the contextual factors suggest that (1) seasonal patterns increase purchase levels but do not significantly affect customer attrition and (2) the first-purchase channel does not significantly affect purchase levels, but the customers acquired online churn faster.

Table 5. Descriptive Statistics for the Multichannel Catalog Merchant Data Set

Three-month cohort	Estimation period	Holdout period	Total
Sample size	—	—	1,402
Available timeframe and split	1 year	6.7 years	7.7 years
Average number of purchases per customer	1.242	2.682	2.089
Standard deviation of repeated purchases	0.619	2.202	2.449
Number of purchases	1741	1188	2929
Zero repeaters	1156	959	850

Table 6. Parameter Estimates for the Multichannel Catalog Merchant Data Set

	Extended Pareto/ NBD	Standard Pareto/ NBD	Description
LL value	−2,025.322	−2,056.098	
BIC	4,123.100	4,141.178	
r	1.548*	1.137*	Homogeneity (purchase process)
α	174.482*	108.174**	Scale parameter (purchase process)
s	0.531	0.181***	Homogeneity (attrition process)
β	8.954	0.373	Scale parameter (attrition)
$\gamma_{purch,1}$	0.194	—	Direct marketing (purchase process)
$\gamma_{purch,2}$	0.821***	—	Seasonality (purchase process)
$\gamma_{purch,3}$	0.210	—	Channel (purchase process; online = 1, offline = 0)
$\gamma_{attr,1}$	−5.131	—	Direct marketing (attrition process)
$\gamma_{attr,2}$	−0.132	—	Seasonality (attrition process)
$\gamma_{attr,3}$	1.907**	—	Channel (attrition process; online = 1, offline = 0)

$^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Table 7 compares the aggregate and individual out-of-sample performance of the extended Pareto/NBD model with that of the standard Pareto/NBD model, the TAM, the Pareto/GGG model, the GPPM, and the LAMDMA based on the future level of transactions (conditional expectation). The extended Pareto/NBD model outperforms all other models at the individual and aggregate level except for the aggregate one-year prediction of the GPPM. The TAM and LAMDMA do not fulfill the Karush-Kuhn-Tucker

Figure 2. In- and Out-Sample Performance for the Multichannel Catalog Merchant Data Set

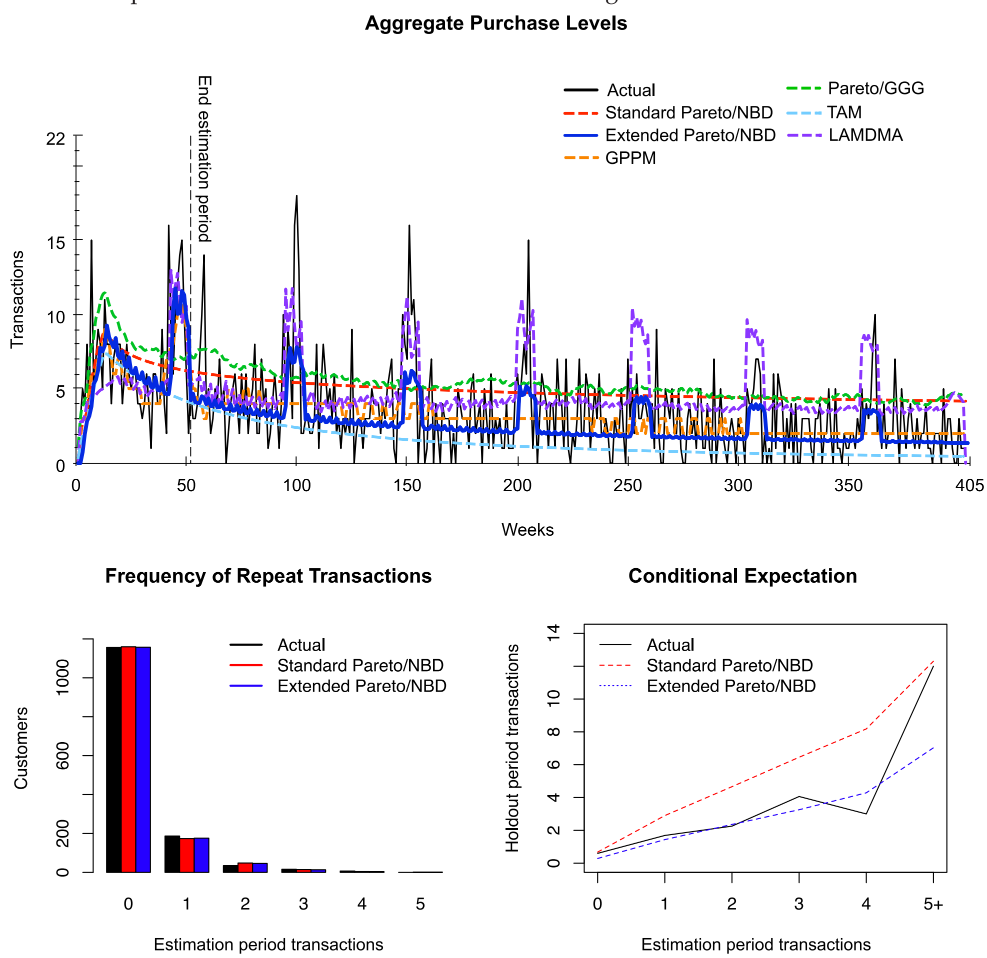


Table 7. Predictive Performance for the Multichannel Catalog Merchant Data Set

	Metric	Extended Pareto/NBD	Standard Pareto/NBD	TAM	Pareto/GGG	GPPM	LAMDMA
One-year prediction	Individual						
	MAE	0.267	0.305	0.328	0.306	0.335	0.301
	Correlation	0.273	0.260	0.270	0.260	0.247	0.241
	Aggregated						
	MAE	2.255	2.783	2.864	3.225	1.759	2.564
Two-year prediction	Individual						
	MAE	0.430	0.513	0.512	0.506	0.598	0.502
	Correlation	0.301	0.275	0.290	0.274	0.292	0.263
	Aggregated						
	MAE	2.168	2.739	2.890	3.005	3.593	2.459
Three-year prediction	Individual						
	MAE	0.551	0.681	0.641	0.664	0.820	0.657
	Correlation	0.345	0.312	0.319	0.311	0.299	0.255
	Aggregated						
	MAE	1.982	2.628	2.685	2.930	3.936	2.431
Four-year prediction	Individual						
	MAE	0.668	0.851	0.763	0.821	1.062	0.813
	Correlation	0.352	0.314	0.320	0.312	0.300	0.250
	Aggregated						
	MAE	1.889	2.447	2.608	2.709	4.255	2.388
Five-year prediction	Individual						
	MAE	0.779	1.024	0.873	0.975	1.347	0.974
	Correlation	0.338	0.292	0.311	0.296	0.277	0.244
	Aggregated						
	MAE	1.812	2.386	2.535	2.624	4.255	2.363
Six-year prediction	Individual						
	MAE	0.874	1.171	0.964	1.108	1.678	1.121
	Correlation	0.331	0.293	0.300	0.291	0.278	0.238
	Aggregated						
	MAE	1.756	2.378	2.462	2.557	4.255	2.384
6.7-year prediction	Individual						
	MAE	0.934	1.275	1.023	1.200	1.951	1.204
	Correlation	0.343	0.282	0.292	0.277	0.268	0.256
	Aggregated						
	MAE	1.704	2.359	2.385	2.517	4.255	2.324

criteria (Kuhn and Tucker 1951). For the maximum length of the holdout period (i.e., 6.7 years), we observe an MAE at the individual level of 0.934 for the extended Pareto/NBD model, 1.275 for the standard Pareto/NDB model, 1.200 for the Pareto/GGG model, 1.014 for the TAM, 1.951 for the GPPM, and 1.204 for the LAMDMA. The improvements in prediction error of the extended Pareto/NBD model are significant. The differences are more than 36% compared with that of the standard Pareto/NBD model, approximately 28% compared with that of the Pareto/GGG model and approximately 13% compared that of the TAM. The results are consistent across all prediction periods. In Figure 2, a comparison of the predictive performance for aggregated purchase levels over time for all models is presented. Notably, the extended Pareto/NBD model is capable of modeling the fluctuations in purchase levels caused by seasonality patterns. Thus, the level of transactions is no longer overestimated during the low season and underestimated during the high season. This observation is also confirmed by the conditional expectation plot in Figure 2, where, in contrast to the standard Pareto/NBD model, expected transactions are predicted more accurately. Additionally, we find for the aggregate purchase levels that the GPPM is performing excellently at the aggregate level in a prediction period of less than one year.

Tables 8 and 9 present the results for the top-tier and second-tier customers. Because of limited variation of the actual transaction during the holdout period, it is not feasible to exactly define the 10% and 20% deciles. For the top and second tier, the next best well-defined deciles to top 10% and 20% are selected (16.6% and 31.6%). The *high, correctly classified* statistic represents the actual best future customers who have been identified as such by the model. We observe that the extended Pareto/NBD model has a better performance in terms of identifying the best future

Table 8. Predictive Performance for Identifying the Top-Tier Customers in the Multichannel Catalog Merchant Data Set

Top-tier customers (6.7-year prediction)	Extended Pareto/NBD	Standard Pareto/NBD	TAM	Pareto/GGG	GPPM	LAMDMA
High, correctly classified (%)	38.79	36.64	37.07	36.21	36.21	36.64
Low, correctly classified (%)	87.86	87.44	87.52	87.35	87.35	87.44
Overall correctly classified (%)	79.74	79.03	79.17	78.89	78.89	79.03
High, incorrectly classified (%)	12.14	12.56	12.48	12.65	12.65	12.56
Low, incorrectly classified (%)	61.21	63.36	62.93	63.79	63.79	63.36
Overall incorrectly classified (%)	20.26	20.97	20.82	21.11	21.11	20.97

customers compared with all other models, except for the top-tier customers, where the LAMDMA performs equally well. When classifying the second-tier future customers over the maximum length of the holdout period, the extended Pareto/NBD model improves the prediction accuracy by more than 40% compared with the standard Pareto/NBD model. The extended Pareto/NBD model identifies 59.59% of the second-tier customers compared with 40.86% identified by the standard Pareto/NBD model, 42.21% identified by the Pareto/GGG, 45.60% by the TAM, and 42.21% by the GPPM. The LAMDMA performs well, identifying 55.53% of the second-tier customers. The results are similar for prediction periods of less than 6.7 years. A popular application of latent attrition models in combination with a Gamma/Gamma model (Colombo and Jiang 1999, Fader et al. 2005b) is prediction of CLV. To do so, DERT (standard Pareto/NBD model) or DECT (extended Pareto/NBD model) metrics are required. The performance assessment of DECT/DERT yields the same results as an assessment based on the conditional expected transactions. This is not surprising, as these metrics are variations of the same statistical concept. Detailed results for DECT/DERT are reported in Appendix D.

4.2. Electronics Retailer

The second data set was obtained from an online retailer selling a wide variety of electronic goods (Ni et al. 2012). We analyze the transaction details for individual customers who purchased first in the three-month period between January and March 1999. These 829 customers account for 5,007 transactions between January 1999 and November 2004. Table 10 provides the key descriptive statistics. Because of an increased interpurchase time, we fit the models for a two-year estimation period. The predictive performance of all four models is evaluated for the following four years.

We model three contextual factors in the extended Pareto/NBD model: (1) seasonal patterns, (2) customer gender, and (3) customer income. The first contextual factor is time varying, whereas the second and third remain constant over time. Increased purchase levels just before Christmas lead to a seasonal pattern in the transaction history of the retailer. Thus, we include the seasonal pattern as a time-varying contextual factor at the aggregate level. The seasonal pattern variable is modeled as a dummy variable based on historical data and expert knowledge. Furthermore, we include information on customer gender and income at the time of the first purchase. We include these customer characteristics as time-invariant contextual factors in our model. Gender is modeled as a dummy variable, whereas customer income is modeled as a continuous variable.

Our results show that the extended Pareto/NBD model has a better in-sample fit than the standard Pareto/NBD model in terms of LL and BIC. The frequency plot of repeat transactions in Figure 3 indicates a mediocre in-sample fit for both extended Pareto/NBD and the standard Pareto/NBD model. The aggregate purchase level plot in Figure 3 indicates a slightly better in-sample fit for the extended Pareto/NBD model as it is capable to cope with the seasonal pattern. Table 11 shows the parameter estimates for the standard and extended Pareto/NBD models. The coefficients of the contextual factors suggest that (1) seasonal patterns increase the level of purchases

Table 9. Predictive Performance for Identifying the Second-Tier Customers in the Multichannel Catalog Merchant Data Set

Second-tier customers (6.7-year prediction)	Extended Pareto/NBD	Standard Pareto/NBD	TAM	Pareto/GGG	GPPM	LAMDMA
High, correctly classified (%)	59.59	40.86	45.60	40.63	42.21	55.53
Low, correctly classified (%)	81.33	71.53	74.66	72.56	73.31	79.46
Overall correctly classified (%)	74.47	61.84	65.48	62.48	63.48	71.90
High, incorrectly classified (%)	18.67	28.47	25.34	27.42	26.69	20.54
Low, incorrectly classified (%)	40.41	59.14	54.40	59.37	57.79	44.47
Overall incorrectly classified (%)	25.54	38.16	34.52	37.52	36.52	28.10

Table 10. Descriptive Statistics for the Electronics Retailer Data Set

Three-month cohort	Estimation period	Holdout period	Total
Sample size	—	—	829
Available timeframe and split	2 year	4 years[a]	6 years
Average number of purchases per customer	3.090	4.570	6.040
Standard deviation of repeated purchases	2.980	5.023	6.406
Number of purchases	2562	2445	5007
Zero repeaters	276	294	129

[a]Data are available only until end of November in the fourth year.

and (2) customers with higher incomes stay longer with the company.

We observe a better out-of-sample prediction at the individual and aggregate levels with the extended Pareto/NBD model. Table 12 compares the out-of-sample predictive accuracy at the individual and aggregate levels for prediction periods of up to four years. In Figure 3, a comparison of the predictive performance for aggregated purchase levels over time is presented. Again, we must acknowledge the capability of the extended Pareto/NBD model to account for seasonal patterns. We observe that the TAM has difficulty fitting the data during the estimation period. The LAMDMA is not used as benchmark for this data set as it requires information on direct marketing activities. Tables 13 and 14 report the performance in terms of identifying the top-tier and second-tier of future customers for a prediction period of four years. As top and second tier, the next best well-defined deciles to top 10% and 20% are selected (11.9% and 20.9%). For the top-tier customers, the extended Pareto/NBD model, the standard Pareto/NBD,

Figure 3. In- and Out-Sample Performance for the Electronics Retailer Data Set

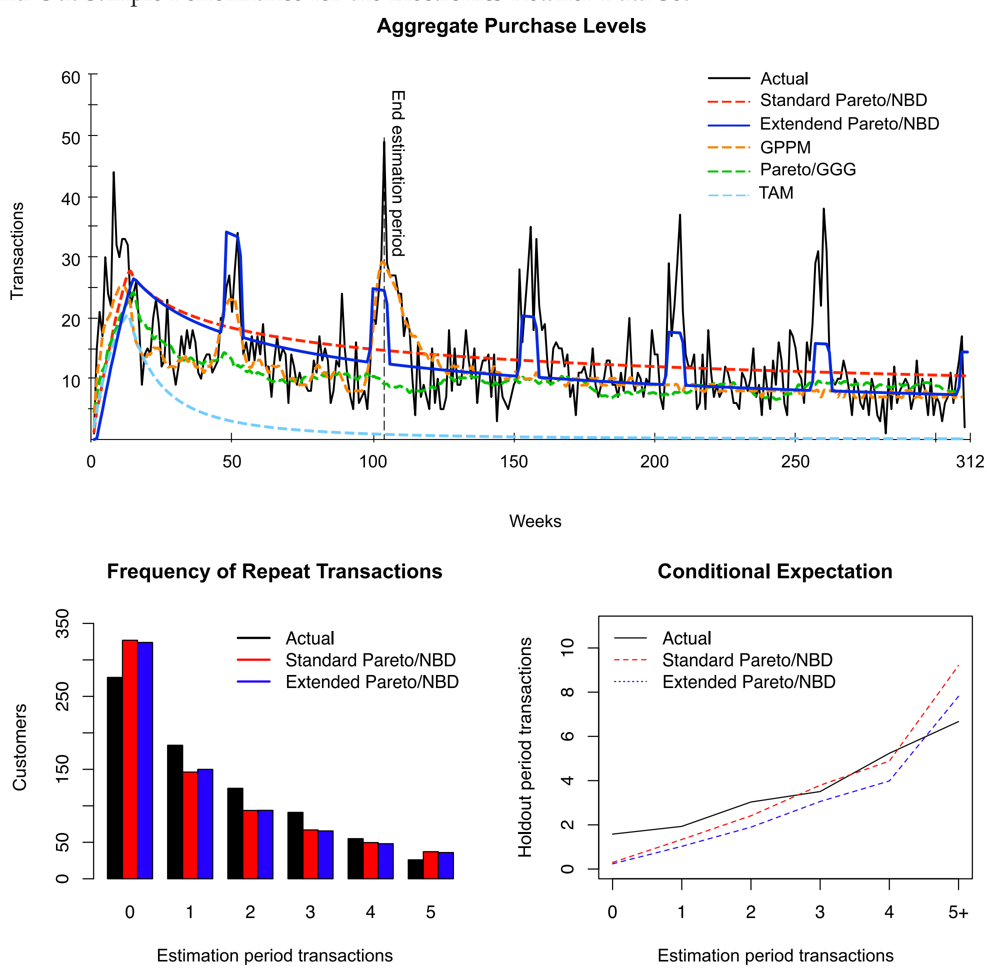


Table 11. Parameter Estimates for the Electronics Retailer Data Set

	Extended Pareto/NBD	Standard Pareto/NBD	Description
LL value	–7,713.724	–7,763.564	
BIC	15,494.651	15,554.009	
r	1.446***	1.633***	Homogeneity (purchase process)
α	36.771***	40.490***	Scale parameter (purchase process)
s	0.524***	0.308***	Homogeneity (attrition process)
β	12.963 ·	7.238**	Scale parameter (attrition)
$\gamma_{purch,1}$	0.674***	—	Seasonality (purchase process)
$\gamma_{purch,2}$	0.021	—	Gender (purchase process; male = 0, female = 1)
$\gamma_{purch,3}$	–0.102	—	Income (purchase process)
$\gamma_{attr,1}$	–0.376	—	Seasonality (attrition process)
$\gamma_{attr,2}$	0.377	—	Gender (attrition process; male = 0, female = 1)
$\gamma_{attr,3}$	–0.978*	—	Income (attrition process)

·; $p < 0.1$; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

and Pareto/GGG models identify 41.41% correctly. The TAM and GPPM reach 40.40%. For the prediction of the second-tier customers, the extended Pareto/NBD model and the standard Pareto/NBD model perform the same with 47.40% and slightly better compared with the Pareto/GGG model (46.24%), the TAM (43.35%), and the GPPM (45.66%). Results for DECT/DERT show similar performance for the standard and the extended Pareto/NDB model. Detailed results for DECT/DERT are reported in Appendix D. We elaborate on possible reasons for varying degrees of performance improvements of the extended Pareto/NBD model in the discussion section.

4.3. Sporting Goods Retailer

The third data set contains transactional records from a sporting goods retailer that cover 1,071 customers who first purchased in the three-month period between January and March 2008. In total, the data set contains 2,226 transactions between January 2008 to December 2013. It is important to note that the retailer is highly specialized and sells goods for a niche sport. The key descriptive statistics are shown in Table 15. Because of the relatively large interpurchase times, the standard and extended Pareto/NBD models are fitted for a two-year estimation period. We evaluate the predictive performance of the models for the following four years.

Table 12. Predictive Performance for the Electronic Retailer Data Set

	Metric	Extended Pareto/NBD	Standard Pareto/NBD	TAM	Pareto/GGG	GPPM
One-year prediction	Individual					
	MAE	0.809	0.835	0.847	0.851	0.976
	Correlation	0.520	0.519	0.401	0.520	0.434
	Aggregated					
	MAE	3.935	5.353	12.164	5.137	7.639
Two-year prediction	Individual					
	MAE	1.493	1.542	1.621	1.541	1.657
	Correlation	0.504	0.504	0.337	0.490	0.419
	Aggregated					
	MAE	4.788	5.941	12.626	5.824	6.919
Three-year prediction	Individual					
	MAE	2.118	2.200	2.353	2.142	2.286
	Correlation	0.491	0.493	0.287	0.481	0.406
	Aggregated					
	MAE	4.573	5.545	12.895	5.807	6.381
Four-year prediction[a]	Individual					
	MAE	2.522	2.659	2.791	2.564	2.783
	Correlation	0.468	0.471	0.248	0.461	0.378
	Aggregated					
	MAE	4.635	5.482	12.738	5.768	6.150

Note. Because no direct marketing variables are available for this data set, LAMDMA results are not reported.
[a]Data are available only until end of November in the fourth year.

Table 13. Predictive Performance in Identifying the Top-Tier Customers in the Electronics Retailer Data Set

Top-tier customers (four-year prediction)	Extended Pareto/NBD	Standard Pareto/NBD	TAM	Pareto/GGG	GPPM
High, correctly classified (%)	42.41	41.41	40.40	41.41	40.40
Low, correctly classified (%)	92.05	92.05	91.92	92.05	91.92
Overall correctly classified (%)	86.01	86.01	85.77	86.01	85.77
High, incorrectly classified (%)	7.95	7.95	8.08	7.95	8.08
Low, incorrectly classified (%)	58.59	58.59	59.60	58.59	59.60
Overall incorrectly classified (%)	13.99	13.99	14.24	13.99	14.23

We include three contextual factors: (1) direct marketing, (2) direct marketing squared, and (3) seasonal buying patterns. The direct marketing activities of the retailer are based on emails, which are sent out by the marketing department. The marketing activities are measured on a weekly basis. There is also an option to opt-out of the newsletters. To model the saturation effects of direct marketing, we also include its squared value (Schroeder and Hruschka 2016). Furthermore, an increased level of purchases is reported by the store manager during the Christmas season. We model this seasonal pattern by using a weekly dummy variable indicating *high season* as an aggregate-level contextual factor (based on expert knowledge). The contextual factor of direct marketing is highly likely to be endogenous. We observe that regular customers are more often subject to marketing emails than customers who do not make frequent purchases. Therefore, we additionally estimate a variation of the extended Pareto/NBD model, where we control for the endogenous direct marketing contextual factors using the proposed copula correction approach.

Our results show that the extended model has a better in-sample fit based on LL and BIC, which is also confirmed by the frequency of repeat transactions plot in Figure 4. We observe that controlling for endogeneity slightly reduces the in-sample fit. All parameter estimates are reported in Table 16. The coefficients of the contextual factors suggest the following. (1) Direct marketing activities increase the transaction levels. However, the impact of direct marketing on customer attrition is not significant. We observe that controlling for endogeneity changes the effect size of direct marketing. (2) Seasonal patterns increase the transaction levels and decrease customer churn.

The extended model outperforms all benchmark models in terms of the out-of-sample prediction and the identification of the future best customers. However, likely because of the small sample size, the GPPM did not converge after multiple trials, and thus, the results are not reported. Table 17 compares the out-of-sample predictive accuracy of the four models at the individual and aggregate levels. The performance is compared for prediction periods of up to four years. We observe better predictive performance across all prediction periods. Controlling for the endogenous contextual direct marketing factor in the estimation and the holdout period reduces the predictive accuracy of the extended Pareto/NBD model. In Figure 4, the model performance for aggregated purchases levels in comparison with the actual levels of transactions are reported. The extended Pareto/NBD model captures the increase in purchase levels caused by seasonal patterns. Tables 18 and 19 illustrate the performance in terms of identifying the top-tier and second-tier future customers for a prediction period of four years. As top and second tier, the next best well-defined deciles to top 10% and 20% are selected (11.67% and 24.18%). We observe that the extended Pareto/NBD model identifies the best future customers more accurately. The extended model identifies 55.77% of the second-tier customers correctly compared with 50.77% in the standard model. This is a 10% improvement. A performance assessment based on DECT/DERT yields the same results. Detailed results are reported in Appendix D.

Table 14. Predictive Performance in Identifying the Second-Tier Customers in the Electronics Retailer Data Set

Second-tier customers (four-year prediction)	Extended Pareto/NBD	Standard Pareto/NBD	TAM	Pareto/GGG	GPPM
High, correctly classified (%)	47.40	47.40	43.35	46.24	45.66
Low, correctly classified (%)	86.13	86.13	85.06	85.82	85.67
Overall correctly classified (%)	78.05	78.05	76.36	77.56	77.32
High, incorrectly classified (%)	13.87	13.87	14.94	14.18	14.33
Low, incorrectly classified (%)	52.60	52.60	56.65	53.76	54.34
Overall incorrectly classified (%)	21.95	21.95	23.64	22.44	22.68

Table 15. Descriptive Statistics for the Sporting Goods Retailer

Three-month cohort	Estimation period	Holdout period	Total
Sample size	—	—	1,071
Available timeframe and split	2 years	4 years	6 years
Average number of purchases per customer	1.490	2.423	2.078
Standard deviation of repeated purchases	0.948	2.580	2.267
Number of purchases	1596	630	2226
Zero repeaters	738	811	629

5. Discussion

Although probabilistic latent customer attrition models are widely used for customer base analyses, it has not been possible to account for time-varying contextual factors in continuous noncontractual settings. Complementing the previous literature, we propose an approach that combines the following properties: (1) the continuous nature of both the purchase and the attrition processes, (2) the inclusion of multiple time-varying and time-invariant contextual factors that can separately influence one, both, or none of the processes, (3) gamma heterogeneity for both processes, (4) the ability to reduce to the standard Pareto/NBD model when it is estimated without any contextual factors, (5) a closed-form maximum-likelihood solution, and (6) the derivation of relevant managerial expressions. These properties have been found to be advantageous by previous studies but have not yet been combined within a single modeling approach.

Using three continuous, noncontractual retailing data sets, we show that including time-invariant and time-varying contextual factors significantly improves the model fit and predictive accuracy over those of the standard Pareto/NBD model and various other

Figure 4. In- and Out-Sample Performance for the Sporting Goods Retailer: The Extended Pareto/NBD Model with and without Copula Results in an Almost Identical Aggregated Curve

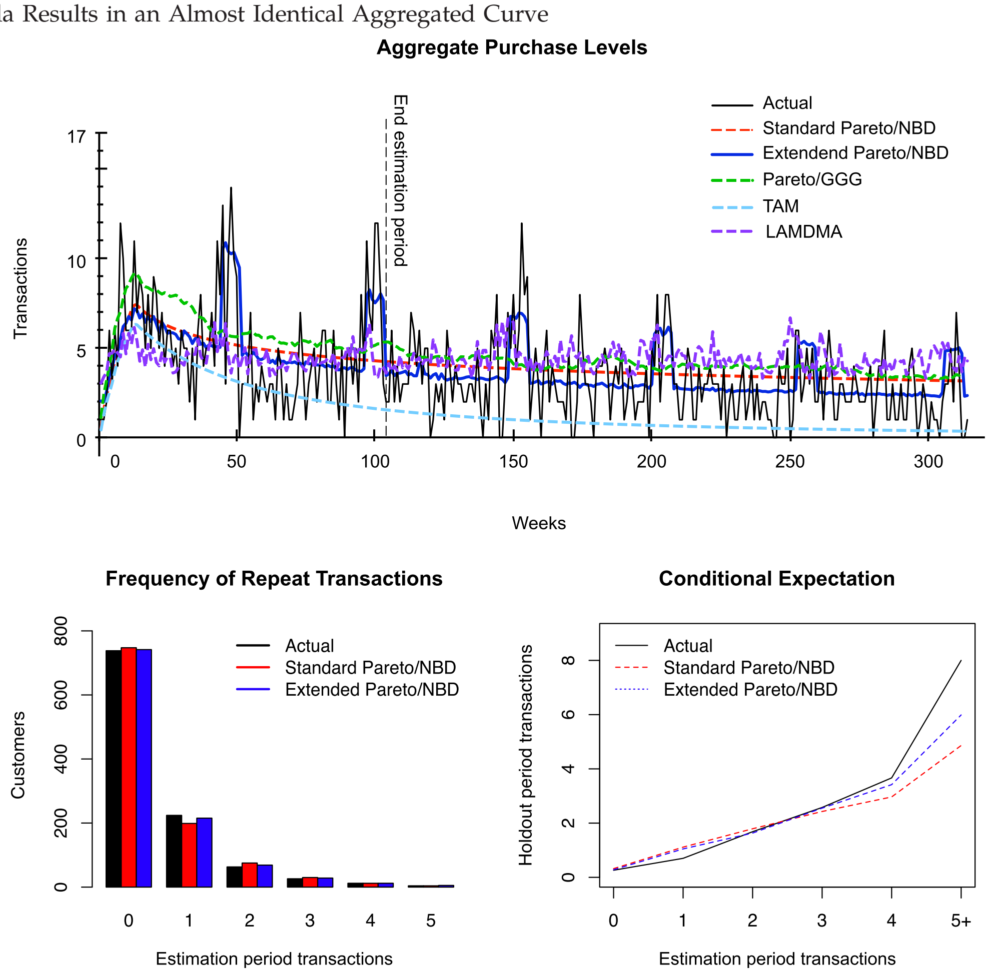


Table 16. Parameter Estimates for the Sporting Goods Retailer Data Set

	Extended Pareto/NBD	Extended Pareto/NBD Copula	Standard Pareto/NBD	Description
LL value	−3,107.808	−3,112.204	−3,212.551	
BIC	6,285.379	6,322.077	6,453.008	
r	1.838***	1.583***	1.412**	Homogeneity (purchase process)
α	179.429**	177.629***	163.862***	Scale parameter (purchase process)
s	0.451**	0.442**	0.270**	Homogeneity (attrition process)
β	29.203*	23.484*	5.784*	Scale parameter (attrition)
$\gamma_{purch,1}$	1.950***	1.911***	—	Direct marketing (purchase process)
$\gamma_{purch,2}$	0.406***	0.336*	—	(Direct marketing)^2 (purchase process)
$\gamma_{purch,3}$	0.747***	0.755***	—	Seasonality (purchase process)
$\gamma_{purch,4}$	—	0.010	—	Copula correction term DM (purchase process)
$\gamma_{purch,5}$	—	0.004	—	Copula correction term DM^2 (purchase process)
$\gamma_{attr,1}$	0.064	0.157	—	Direct marketing (attrition process)
$\gamma_{attr,2}$	2.591	2.251	—	(Direct marketing)^2 (attrition process)
$\gamma_{attr,3}$	−0.101**	−0.091**	—	Seasonality (attrition process)
$\gamma_{attr,4}$	—	−0.013	—	Copula correction term DM (attrition process)
$\gamma_{attr,5}$	—	−0.222	—	Copula correction term DM^2 (attrition process)

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Pareto- and non–Pareto-type models. In particular, we observe better accuracy for individual-level predictions for the first and third data set. For these datasets, we include time-varying contextual factors that affect customers not only at the aggregate level, that is, all customers in the same way, but also at the individual level. For the second data set, we include only an aggregate-level time-varying contextual factor. Although we still find better individual-level predictive accuracy, the improvements are smaller than those of the other two data sets. To significantly improve the individual-level prediction, additional time-varying information on the individual customer is needed. Adding time-varying contextual factors that affect the customers at the aggregate-level only (i.e., seasonality pattern) mainly leads to better aggregate-level predictions; however, they do not seem to have a significant effect on the individual-level predictions. In summary, adding contextual factors enhances the managerial applicability of the Pareto/NBD model because it enables managers to better identify the best future customers and paves the way for more accurate individual predictions for future customer purchase and attrition behavior. Hence, fully leveraging the capabilities of the proposed model requires that time-varying contextual factors be considered. Although this

Table 17. Predictive Performance for the Sporting Goods Retailer Data Set

	Metric	Extended Pareto/NBD	Extended Pareto/NBD Copula	Standard Pareto/NBD	TAM	Pareto/GGG	LAMDMA
One-year prediction	Individual						
	MAE	0.274	0.275	0.547	0.284	0.283	0.335
	Correlation	0.484	0.459	0.286	0.448	0.468	0.143
	Aggregated						
	MAE	1.989	1.991	2.042	2.081	2.225	2.075
Two-year prediction	Individual						
	MAE	0.452	0.452	0.547	0.467	0.470	0.587
	Correlation	0.524	0.500	0.286	0.503	0.519	0.140
	Aggregated						
	MAE	1.824	1.828	1.885	2.704	2.056	2.058
Three-year prediction	Individual						
	MAE	0.594	0.599	0.715	0.626	0.625	0.799
	Correlation	0.551	0.538	0.296	0.518	0.533	0.191
	Aggregated						
	MAE	1.659	1.660	1.759	2.492	1.954	2.045
Four-year prediction	Individual						
	MAE	0.741	0.759	0.843	0.787	0.779	1.010
	Correlation	0.551	0.536	0.297	0.518	0.532	0.198
	Aggregated						
	MAE	1.610	1.624	1.698	2.412	1.877	2.048

Note. The GPPM did not converge; results are not reported.

Table 18. Predictive Performance in Identifying the Top-Tier Customers in the Sporting Goods Retailer Data Set

Top-tier customers (four-year prediction)	Extended Pareto/NBD	Extended Pareto/NBD Copula	Standard Pareto/NBD	TAM	Pareto/GGG	LAMDMA
High, correctly classified (%)	44.00	43.20	43.20	28.00	41.60	41.60
Low, correctly classified (%)	92.60	92.49	92.49	90.47	92.28	92.18
Overall correctly classified (%)	86.93	86.74	86.74	83.19	86.37	86.27
High, incorrectly classified (%)	7.39	7.51	7.51	9.51	7.72	7.82
Low, incorrectly classified (%)	56.00	56.80	56.80	72.00	58.40	58.40
Overall incorrectly classified (%)	13.53	13.26	13.07	16.81	13.63	13.73

Note. The GPPM did not converge; therefore, the results for this model are not reported.

is easily achieved for regular aggregate and individual patterns, it is not always the case for other factors such as direct marketing. However, direct marketing efforts are often predetermined in the short to medium term. For example, various marketing activities are timed along a customer's tenure, and other marketing campaigns are repeated on a regular basis (e.g., monthly newsletters and back-to-school sales). In terms of short-term predictions, direct marketing campaigns are often planned several months in advance. In addition to increasing predictive accuracy, the proposed model can measure the effect sizes of contextual factors affecting the underlying purchase and attrition processes and can disentangle the effects between the two processes. In particular, the proposed model is able to control for endogenous contextual factors and can identify the impact of contextual factors on customers' purchase and attrition processes.

This paper contributes to marketing research in many substantive ways. We introduce time-varying contextual factors to the Pareto/NBD model, which is widely used in research and practice and is considered the *gold standard* in the field (Jerath et al. 2011, Singh and Jain 2013). This is the first time that a continuous noncontractual model incorporates both time-varying and time-invariant contextual factors. In particular, this paper complements the existing practices used for customer base analyses in five distinctive ways: (1) *Increasing predictive accuracy.* Including time-varying contextual factors significantly increases the predictive accuracy for future customer behavior at both the individual and aggregated levels. Practitioners and researchers can explicitly consider, for example, seasonalities and nonrandom firm activities, such as prescheduled direct and mass marketing campaigns, when predicting future transaction levels. Because these activities directly aim to influence customer purchase behavior, it is crucial to take them into account. In this context, we observe differences in the increase in predictive accuracy depending on the scope of the contextual factors; in other words, individual-level contextual factors increase predictive accuracy more than aggregate-level contextual factors. (2) *Identifying the determinants of customers' purchase and attrition processes.* The proposed model can reliably identify the impact of exogenous contextual factors of the two underlying processes. Thereby, we maintain the capability of the standard Pareto/NBD model to simultaneously model customers' purchase and attrition behavior. Thus, the proposed model allows us to disentangle the impact of the contextual factors of both processes. Additionally, we propose three generalized approaches to control for endogeneity arising in both the purchase and attrition processes. This allows us to reliably identify the impact of endogenous contextual factors. To clarify, we can answer the following question: do certain marketing activities increase the number of transactions, but lead to a higher churn at the same time? However, endogeneity is of lesser importance when the focus is primarily on predictive accuracy. (3) *Using a closed-form expression.* The closed-form solution is a key feature of the standard Pareto/NBD model and is one reason for its success. The proposed model retains this

Table 19. Predictive Performance in Identifying the Second-Tier Customers in the Sporting Goods Retailer Data Set

Second-tier customers (four-year prediction)	Extended Pareto/NBD	Extended Pareto/NBD Copula	Standard Pareto/NBD	TAM	Pareto/GGG	LAMDMA
High, correctly classified (%)	55.77	50.57	50.77	41.31	50.77	51.35
Low, correctly classified (%)	85.78	84.24	84.22	81.16	84.22	84.48
Overall correctly classified (%)	78.34	76.10	76.10	71.52	76.10	76.47
High, incorrectly classified (%)	14.22	15.76	15.78	18.84	15.78	15.52
Low, incorrectly classified (%)	44.23	49.42	49.23	58.69	49.23	48.65
Overall incorrectly classified (%)	21.47	23.90	23.90	28.48	23.90	23.53

Note. The GPPM did not converge; therefore, the results for this model are not reported.

key feature. The closed-form solution allows us to derive additional expressions such as P(alive) and the conditional expectation for any customer. These expressions allow researchers and practitioners to gain an intuitive understanding of the predicted customer behavior. Additionally, a closed-form solution allows for robust, scalable, and efficient implementation. Although the computational resources needed for calculations have increased significantly during the last few years, in an ever-increasing competitive environment, efficient implementation is still key for marketers to gain access to relevant and reliable information in a timely manner. (4) *Scenario analysis.* By proposing an approach that involves an efficient model that controls for endogenous contextual factors, we have taken the first step toward enabling practitioners to perform scenario analyses (Hruschka 2010). Given that many contextual factors can be directly influenced by the company itself, a scenario analysis might help practitioners make the right decisions. For example, in the case of direct marketing activities, managers, who have access to information on the cost of specific actions and are bound to budget restrictions, could assess a hypothetical sequence of various personalized marketing campaigns to determine which one is most efficient in improving either the lifetime value of an individual customer or the firm's customer equity. (5) *DECT.* To ensure the managerial relevance of modeling customer latent attrition, we propose, similar to Fader et al. (2005b), a corresponding metric. The proposed DECT metric measures the flow of future transactions discounted to the present value, which is important for managerial applications. Furthermore, managers can assess individual customer valuations by using the DECT metric in combination with a Gamma/Gamma model for modeling customer spending.

Finally, we identified five limitations that can be addressed by future research efforts in this field. First, the impact of tactical marketing activities is usually not restricted to the customer purchase and attrition processes. These marketing activities likely influence customer spending. Neither the Pareto/NBD model nor our extension accounts for the spending process of customers. Further research could focus on modeling the impact of contextual factors on the spending process. Second, a study may leverage the extended Pareto/NBD model to gain further insights on the importance of clumpiness (Zhang et al. 2015) or differences in the interpurchase times of customers (Platzer and Reutterer 2016). Platzer and Reutterer (2016) showed that patterns of interpurchase times may provide additional insights for the prediction of future customer behavior. However, their model does not account for any contextual factors other than fixed regularity patterns. Future research may combine the capabilities of the extended Pareto/NBD and the Pareto/GGG model. Third, future research may analyze the individual responsiveness of customers to contextual factors. The proposed extended Pareto/NBD model assumes that all customers have the same degree of responsiveness. A fourth promising direction for future research is to model common patterns across cohorts. Most studies on probabilistic customer attrition models, including this one, assume that the cohorts are independent of each other. Although it is possible to use a pooled model and to account for cohort-specific differences by including fixed effects in the extended Pareto/NBD model, this is not an optimal solution. Leveraging the commonalities between cohorts, similar to the cross-cohort model introduced by Gopalakrishnan et al. (2017), could possibly improve the predictions, especially for cohorts that have been observed for only a relatively short period of time. Fifth, future research may add further benchmarks. Although we provide a comparison with a wide set of benchmark models, there exist further approaches that the proposed extended Pareto/NBD model can be compared with (e.g., the hidden Markow model proposed by Netzer et al. 2008 or the approach proposed by Allenby et al. 1999). Finally, on a different note, the computational aspects of the extended Pareto/NBD model deserve further attention. The complexity of the extended Pareto/NBD model increases computational requirements. See Table 20 for a run time comparison of all benchmark

Table 20. Run Time Comparison

	Model	Timing	Implementation
With time-varying contextual factors	Extended Pareto/NBD	6 h, 51 min	CLVTools (Bachmann et al. 2020)
	Transaction attribute	8 h, 10 min	Own
	GPPM	9 h, 57 min	Dew and Ansari (2018)
	LAMDMA	5 h, 48 min	Own
No time-varying contextual factors	Standard Pareto/NBD	<1 min	CLVTools (Bachmann et al. 2020)
	Pareto/GGG	8 min	BTYDplus (Platzer 2018)

Note. Detailed specifications for the run time comparison are provided in Appendix E.

models used in this paper. Although such benchmarks are always dependent on the underlying implementation and are subject to a natural improvement with the continuous evolution of computational hardware, it is obvious that applied researchers or practitioners may focus on improving the computational efficiency of the extended Pareto/NBD model.

Acknowledgments

The authors are listed in alphabetical order and contributed equally.

Appendix A. Closed-Form Solutions

We use the definitions and symbols defined in Table A.1.

A.1. Closed-Form Likelihood Expression

Following the specification in Table A.1, we obtain the following closed-form solution for the likelihood. Additional derivations of the expression are provided in Appendix B.

$$L(\alpha, r, \beta, s, \gamma_{purch}, \gamma_{attr} | \mathbf{X}^P, \mathbf{X}^A, x, \mathbf{t}, T) = \left(\prod_{j=1}^{x} A_{k_j}\right) \frac{\Gamma(x+r)\alpha^r \beta^s}{\Gamma(r)} \left[\frac{s}{r+s+x}\mathbb{Z} + \frac{1}{\left(\sum_{j=1}^{x} B_{k_j} + B_{k_T} + \alpha\right)^{x+r} \left(D_{k_T} + \beta\right)^s}\right], \tag{A.1}$$

where, for $k_T > 1$

$$\mathbb{Z} = \mathbb{Y}_1 + \mathbb{Y}_{k_T} + \sum_{i=2}^{k_T - 1} \mathbb{Y}_i,$$

with

$$\begin{aligned}
\mathbb{Y}_1 = {} & \mathbb{I}\{a_1 + \alpha > (b_1 + \beta)A_1/C_1\} \times \left(\frac{A_1}{C_1}\right)^s \\
& \times \left(\frac{1}{(a_1 + (1-d_1)A_1 + \alpha)^{r+s+x}}\right. \\
& \times {}_2F_1\left(r+s+x, s+1; r+s+x+1;\right. \\
& \left.\times \frac{a_1 + \alpha - (b_1+\beta)A_1/C_1}{a_1 + (1-d_1)A_1 + \alpha}\right) - \frac{1}{(a_1 + A_1 + \alpha)^{r+s+x}} \\
& \left.{}_2F_1\left(r+s+x, s+1; r+s+x+1;\ \frac{a_1 + \alpha - (b_1+\beta)A_1/C_1}{a_1 + A_1 + \alpha}\right)\right) \\
& + \mathbb{I}\{a_1 + \alpha < (b_1 + \beta)A_1/C_1\} \times \left(\frac{A_1}{C_1}\right)^s \\
& \times \left(\frac{1}{((b_1 + (1-d_1)C_1 + \beta)A_1/C_1)^{r+s+x}}\right. \\
& {}_2F_1\left(r+s+x, r+x; r+s+x+1; \frac{(b_1+\beta)A_1/C_1 - (a_1+\alpha)}{(b_1 + (1-d_1)C_1 + \beta)A_1/C_1}\right) \\
& - \frac{1}{((b_1 + C_1 + \beta)A_1/C_1)^{r+s+x}} \cdot \\
& \left.{}_2F_1\left(r+s+x, r+x; r+s+x+1;\ \frac{(b_1+\beta)A_1/C_1 - (a_1+\alpha)}{(b_1 + C_1 + \beta)A_1/C_1}\right)\right),
\end{aligned}$$

Table A.1. Definitions of Variables Used in This Article

Variables	Definition
t_j, t_x	Time/date of the jth and xth purchase
$z_{j-1,j}$	Time between purchase $(j-1)$ and purchase j
k_j	Number of intervals between purchase $(j-1)$ and purchase j, denoting the interval including transaction j as 1
ω	Time of the death of the customer
T	Time of the end of the observation period
$k_{x,\omega}$	Number of intervals between purchase x and ω including the first and the last
$k_T = k_{x,T}$	Number of intervals between purchase x and T including the first and the last
K	Total number of intervals: $[= \sum_{j=1}^{x} k_j + k_T - x]$
$\mathbf{x}_k^P$	Vector of the contextual factors of the purchase process with values in the kth interval
$\mathbf{x}_k^A$	Vector of the contextual factors of the attrition process with values in the kth interval
$\mathbf{X}^P$	Matrix of the contextual factors of the purchase process over all K intervals: $[= (\mathbf{x}_1^P \mathbf{x}_2^P \cdots \mathbf{x}_K^P)']$
$\mathbf{X}^A$	Matrix of the contextual factors of the attrition process over all K intervals: $[= (\mathbf{x}_1^A \mathbf{x}_2^A \cdots \mathbf{x}_K^A)']$
γ	Vector of contextual effects (purchase and attrition).
d_T	Time difference between T and the end of the interval T is contained in
d_1	Time of the transaction $(j-1)$ to the end of the first contextual factor interval for any two successive transactions $(j-1)$ and j. Note that d_1 is indexed relative to the purchases
d_2	Time difference between the initial purchase and the end of the interval the initial purchase is contained in
δ	$\begin{cases} 1 \text{ if } k_j \geq 2 \\ 0 \text{ if } k_j = 1 \end{cases}$
$\mathbf{t}$	$(t_1, \ldots, t_x)$
Functions	
${}_2F_1(a, b; c; z)$	Gaussian hypergeometric function
$\Gamma(x)$	Gamma function
Ψ	Confluent hypergeometric function of the second kind

and

$$\begin{aligned}
\mathbb{Y}_i = {} & \mathbb{I}\{a_i + \alpha > (b_i + \beta)A_i/C_i\} \times \left(\frac{A_i}{C_i}\right)^s \left(\frac{1}{(a_i + \alpha)^{r+s+x}}\right. \\
& \times {}_2F_1\left(r+s+x, s+1; r+s+x+1;\ \frac{a_i + \alpha - (b_i+\beta)A_i/C_i}{a_i + \alpha}\right) \\
& - \frac{1}{(a_i + A_i + \alpha)^{r+s+x}} {}_2F_1\left(r+s+x, s+1; r+s+x+1;\right. \\
& \left.\left.\times \frac{a_i + \alpha - (b_i+\beta)A_i/C_i}{a_i + A_i + \alpha}\right)\right) + \mathbb{I}\{a_i + \alpha < (b_i + \beta)A_i/C_i\} \times \left(\frac{A_i}{C_i}\right)^s \\
& \times \left(\frac{1}{((b_i + \beta)A_i/C_i)^{r+s+x}}\right.\ {}_2F_1\left(r+s+x, r+x; r+s+x+1;\right. \\
& \left.\times \frac{(b_i+\beta)A_i/C_i - (a_i+\alpha)}{(b_i+\beta)A_i/C_i}\right) - \frac{1}{((b_i + C_i + \beta)A_i/C_i)^{r+s+x}} \\
& \left.{}_2F_1\left(r+s+x, r+x; r+s+x+1;\ \frac{(b_i+\beta)A_i/C_i - (a_i+\alpha)}{(b_i + C_i + \beta)A_i/C_i}\right)\right),
\end{aligned}$$

for $i = 2, \dots, k-1$ and

$$\begin{aligned}
\mathbb{Y}_{k_T} &= \mathbb{I}\{a_{k_T} + \alpha > (b_{k_T} + \beta)A_{k_T}/C_{k_T}\} \times \left(\frac{A_{k_T}}{C_{k_T}}\right)^s \\
&\times \left(\frac{1}{(a_{k_T} + \alpha)^{r+s+x}} \; {}_2F_1\left(r+s+x, s+1; r+s+x+1; \right.\right. \\
&\left. \frac{a_{k_T} + \alpha - (b_{k_T} + \beta)A_{k_T}/C_{k_T}}{a_{k_T} + \alpha}\right) - \frac{1}{(a_T^* + \alpha)^{r+s+x}} \\
&{}_2F_1\left(r+s+x, s+1; r+s+x+1; \right. \\
&\left.\left. \frac{a_{k_T} + \alpha - (b_{k_T} + \beta)A_{k_T}/C_{k_T}}{a_T^* + \alpha}\right)\right) \\
&+ \mathbb{I}\{a_{k_T} + \alpha < (b_{k_T} + \beta)A_{k_T}/C_{k_T}\} \times \left(\frac{A_{k_T}}{C_{k_T}}\right)^s \\
&\times \left(\frac{1}{((b_{k_T} + \beta)A_{k_T}/C_{k_T})^{r+s+x}}\right. \\
&{}_2F_1\left(r+s+x, r+x; r+s+x+1; \right. \\
&\left. \frac{(b_{k_T} + \beta)A_{k_T}/C_{k_T} - (a_{k_T} + \alpha)}{(b_{k_T} + \beta)A_{k_T}/C_{k_T}}\right) - \frac{1}{((b_T^* + \beta)A_{k_T}/C_{k_T})^{r+s+x}} \\
&{}_2F_1\left(r+s+x, r+x; r+s+x+1; \right. \\
&\left.\left. \frac{(b_{k_T} + \beta)A_{k_T}/C_{k_T} - (a_{k_T} + \alpha)}{(b_T^* + \beta)A_{k_T}/C_{k_T}}\right)\right).
\end{aligned}$$

While for $k_T = 1$,

$$\begin{aligned}
\mathbb{Z} &= \mathbb{I}\{a_1 + \alpha > (b_1 + \beta)A_1/C_1\} \times \left(\frac{A_1}{C_1}\right)^s \\
&\times \left(\frac{1}{(a_1 + (1-d_1)A_1 + \alpha)^{r+s+x}}\right. \\
&{}_2F_1\left(r+s+x, s+1; r+s+x+1; \; \frac{a_1 + \alpha - (b_1 + \beta)A_1/C_1}{a_1 + (1-d_1)A_1 + \alpha}\right) \\
&- \frac{1}{(a_1^T + \alpha)^{r+s+x}} \; {}_2F_1\left(r+s+x, s+1; r+s+x+1; \right. \\
&\left.\left. \frac{a_1 + \alpha - (b_1 + \beta)A_1/C_1}{a_1^T + \alpha}\right)\right) \\
&+ \mathbb{I}\{a_1 + \alpha < (b_1 + \beta)A_1/C_1\} \times \left(\frac{A_1}{C_1}\right)^s \\
&\times \left(\frac{1}{((b_1 + (1-d_1)C_1 + \beta)A_1/C_1)^{r+s+x}}\right. \\
&{}_2F_1\left(r+s+x, r+x; r+s+x+1; . \right. \\
&\left. \frac{(b_1 + \beta)A_1/C_1 - (a_1 + \alpha)}{(b_1 + (1-d_1)C_1 + \beta)A_1/C_1}\right) - \frac{1}{((b_1^T + \beta)A_1/C_1)^{r+s+x}} \\
&{}_2F_1\left(r+s+x, s+1; r+s+x+1; . \right. \\
&\left.\left. \frac{(b_1 + \beta)A_1/C_1 - (a_1 + \alpha)}{(b_1^T + \beta)A_1/C_1}\right)\right),
\end{aligned}$$

with $i = 1, \dots, k_T$,

$$\begin{aligned}
A_{k_j} &= \exp\left(\gamma'_{purch}\mathbf{x}^P_{k_j}\right), \\
B_{k_j} &= \exp\left(\gamma'_{purch}\mathbf{x}^P_1\right)d_1 + \sum_{l=2}^{k_j-1} \exp\left(\gamma'_{purch}\mathbf{x}^P_l\right) \\
&\quad + \exp\left(\gamma'_{purch}\mathbf{x}^P_{k_j}\right)[z_{j-1,j} - d_1 - \delta(k_j - 2)], \\
A_i &= \exp\left(\gamma'_{purch}\mathbf{x}^P_i\right), \\
C_i &= \exp(\gamma'_{attr}\mathbf{x}^A_i), \\
\bar{B}_i &= \exp\left(\gamma'_{purch}\mathbf{x}^P_1\right)d_1 + \sum_{l=2}^{i-1} \exp(\gamma_{purch}\mathbf{x}^P_l) \\
&\quad + \exp\left(\gamma'_{purch}\mathbf{x}^P_i\right)[-t_x - d_1 - \mathbb{I}\{i \geq 2\}(i-2)], \\
\bar{D}_i &= \exp(\gamma'_{attr}\mathbf{x}^A_1)d_2 + \sum_{l=2}^{k_{0,x}+i-2} \exp(\gamma'_{attr}\mathbf{x}^A_l) + \\
&\quad + \exp\left(\gamma'_{attr}\mathbf{x}^A_{k_{0,x}+i-1}\right)\left[-d_2 - \mathbb{I}\{k_{0,x} + i - 1 \geq 2\}\right. \\
&\quad \times (k_{0,x} + i - 3)], \\
D_{k_T} &= \exp\left(\gamma'_{attr}\mathbf{x}^A_{k_{0,x}+k_T-1}\right)T + \bar{D}_{k_T}, \\
a_i &= \sum_{j=1}^{x} B_{k_j} + \bar{B}_i + (t_x + d_1 + (i-2))A_i, \\
b_i &= \bar{D}_i + (t_x + d_1 + (i-2))C_i, \\
a_T^* &= \sum_{j=1}^{x} B_{k_j} + \bar{B}_{k_T} + TA_{k_T}, \\
b_T^* &= \bar{D}_{k_T} + TC_{k_T}, \\
a_1^T &= \sum_{j=1}^{x} B_{k_j} + \bar{B}_1 + TA_1, \\
b_1^T &= \bar{D}_1 + TC_1.
\end{aligned}$$

A.2. P(alive)

Following the specifications in Table A.1 and Appendix A.1, we obtain the following closed-form solution for P(alive):

$$\begin{aligned}
&P(\Omega > T | r, \alpha, s, \beta, \gamma_{purch}, \gamma_{attr}, \mathbf{X}^P, \mathbf{X}^A, x, t_x, T) \\
&= \prod_{j=1}^{x} A_{k_j} \frac{\Gamma(x+r)\alpha^r \beta^s}{\Gamma(r)\left(\sum_{j=1}^{x} B_{k_j} + B_{k_T} + \alpha\right)^{x+r} (D_{k_T} + \beta)^s} \\
&\quad \times \frac{1}{L(\alpha, r, \beta, s, \gamma_{purch}, \gamma_{attr} | \mathbf{X}^P, \mathbf{X}^A, x, \mathbf{t}, T)} \\
&= \int_0^\infty \int_0^\infty \left[\prod_{j=1}^{x} \lambda_0 A_{k_j} \exp\left[-\lambda_0 B_{k_j}\right]\right] \exp[-\lambda_0 B_{k_T}] \\
&\quad \times \exp[-\mu_0 D_{k_T}] \; g(\lambda_0) g(\mu_0) \; d\lambda_0 d\mu_0 \\
&\quad \times \frac{1}{L(\alpha, r, \beta, s, \gamma_{purch}, \gamma_{attr} | \mathbf{X}^P, \mathbf{X}^A, x, \mathbf{t}, T)} \\
&= \frac{1}{\left(\sum_{j=1}^{x} B_{k_j} + B_{k_T} + \alpha\right)^{x+r} (D_{k_T} + \beta)^s \frac{s}{r+s+x} \mathbb{Z} + 1}. \qquad \text{(A.2)}
\end{aligned}$$

A.3. Conditional Expectation

In the following, we present the closed-form solution for the conditional expectation. Additional derivations of the expression are provided in the Appendix C.

Following specification in Table A.1, we obtain the following closed-form solution for the conditional expectation in the case of $k_{T,T+t} \geq 2$:

$$\begin{aligned}
&E[Y(T,T+t)\,|\,r,\alpha,s,\beta,\gamma_{purch},\gamma_{attr},\mathbf{X}^P,\mathbf{X}^A,x,\mathbf{t},T]\\
&= \frac{\prod_{j=1}^{x} A_{k_j}}{L(\alpha,r,\beta,s,\gamma_{purch},\gamma_{attr}|\mathbf{X}^P,\mathbf{X}^A,x,\mathbf{t},T)}\\
&\quad\times \frac{\Gamma(x+r)\alpha^r\beta^s}{\Gamma(r)\left(\sum_{j=1}^{x} B_{k_j}+B_{k_T}+\alpha\right)^{x+r}(D_{k_T}+\beta)^s}\\
&\quad\times \frac{(r+x)(D_{k_T}+\beta)^s}{\left(\sum_{j=1}^{x} B_{k_j}+B_{k_T}+\alpha\right)(s-1)}\left[\frac{B_{k_{T,T+t}}(s-1)}{(D_{k_{T,T+t}}+\beta)^s}\right.\\
&\quad\left.+S_1^*+S^*_{k_{T,T+t}}+\mathbb{I}_{\{3,4,\ldots\}}(k_{T,T+t})\sum_{i=2}^{k_{T,T+t}-1} S_i\right],
\end{aligned} \tag{A.3}$$

and for $k_{T,T+t}=1$ we obtain:

$$\begin{aligned}
&E[Y(T,T+t)|r,\alpha,s,\beta,\gamma_{purch},\gamma_{attr},\mathbf{X}^P,\mathbf{X}^A,x,\mathbf{t},T]\\
&= \frac{\prod_{j=1}^{x} A_{k_j}}{L(\alpha,r,\beta,s,\gamma_{purch},\gamma_{attr}|\mathbf{X}^P,\mathbf{X}^A,x,\mathbf{t},T)}\frac{\alpha^r\beta^s}{\Gamma(r)(s-1)}\\
&\quad\times \frac{\Gamma(x+r+1)}{\left(\sum_{j=1}^{x} B_{k_j}+B_{k_T}+\alpha\right)^{x+r+1}}\left[\frac{B_{k_{T,T+t}}(s-1)}{(D_{k_{T,T+t}}+\beta)^s}\right.\\
&\quad+\frac{A_1[Ts+1/C_1(\bar{D}_1+\beta)]+\bar{B}_1(s-1)}{(\bar{D}_1+\beta+C_1T)^s}\\
&\quad\left.-\frac{A_1[(T+t)s+1/C_1(\bar{D}_1+\beta)]+\bar{B}_1(s-1)}{(\bar{D}_1+\beta+C_1(T+t))^s}\right],
\end{aligned} \tag{A.4}$$

with $i=1,\ldots,k_{T,T+t}$,

$$\begin{aligned}
S_i &= \frac{A_i[b_i^T s+1/C_i(\bar{D}_i+\beta)]+\bar{B}_i(s-1)}{(\bar{D}_i+\beta+C_i b_i^T)^s}\\
&\quad-\frac{A_i[(b_i^T+1)s+1/C_i(\bar{D}_i+\beta)]+\bar{B}_i(s-1)}{(\bar{D}_i+\beta+C_i(b_i^T+1))^s},\\
S_1^* &= \frac{A_1[Ts+1/C_1(\bar{D}_1+\beta)]+\bar{B}_1(s-1)}{(\bar{D}_1+\beta+C_1T)^s}\\
&\quad-\frac{A_1[(T+d_T)s+1/C_1(\bar{D}_1+\beta)]+\bar{B}_1(s-1)}{(\bar{D}_1+\beta+C_1(T+d_T))^s},\\
S^*_{k_{T,T+t}} &= \frac{A_{k_{T,T+t}}\left[b^T_{k_{T,T+t}}s+1/C_i(\bar{D}_{k_{T,T+t}}+\beta)\right]+\bar{B}_{k_{T,T+t}}(s-1)}{\left(\bar{D}_{k_{T,T+t}}+\beta+C_{k_{T,T+t}}b^T_{k_{T,T+t}}\right)^s}\\
&\quad-\frac{A_{k_{T,T+t}}[(T+t)s+1/C_{k_{T,T+t}}(\bar{D}_{k_{T,T+t}}+\beta)]+\bar{B}_{k_{T,T+t}}(s-1)}{(\bar{D}_{k_{T,T+t}}+\beta+C_{k_{T,T+t}}(T+t))^s},
\end{aligned}$$

$$\begin{aligned}
B_{k_{T,T+t}} &= A_{k_{T,T+t}}(T+t)+\bar{B}_{k_{T,T+t}},\\
D_{k_{T,T+t}} &= C_{k_{T,T+t}}(T+t)+\bar{D}_{k_{T,T+t}},
\end{aligned}$$

$$\begin{aligned}
\bar{B}_i &= \exp\left(\gamma'_{purch}\mathbf{x}_1^P\right)d_T+\sum_{l=2}^{i-1}\exp\left(\gamma'_{purch}\mathbf{x}_l^P\right)\\
&\quad+\exp\left(\gamma'_{purch}\mathbf{x}_i^P\right)[-T-d_T-\delta(i-2)],\\
\bar{D}_i &= \exp(\gamma'_{attr}\mathbf{x}_1^A)d_2+\sum_{l=2}^{k_{0,T}+i-2}\exp(\gamma'_{attr}\mathbf{x}_l^A)\\
&\quad+\exp\left(\gamma'_{attr}\mathbf{x}^A_{k_{0,T}+i-1}\right)[-d_2-\delta(k_{0,T}+i-3)],\\
A_i &= \exp\left(\gamma'_{purch}\mathbf{x}_i^P\right),\\
C_i &= \exp\left(\gamma'_{attr}\mathbf{x}^A_{k_{0,T}+i-1}\right),\\
B_{k_j} &= \exp\left(\gamma'_{purch}\mathbf{x}_1^P\right)d_T+\sum_{l=2}^{k_j-1}\exp\left(\gamma'_{purch}\mathbf{x}_l^P\right)\\
&\quad+\exp\left(\gamma'_{purch}\mathbf{x}^P_{k_j}\right)[z_{j-1,j}-d_T-\delta(k_j-2)],\\
b_i^T &= T+d_T+(i-2),\text{ with } i=2,..,k_{T,T+t}.
\end{aligned}$$

A.4. DECT

In this section, we obtain the following expression for DECT by solving by removing the conditioning on λ_0 and μ_0 in (26). See Table A.1 for specifications. A step-by-step derivation of the expression is provided in Online Appendix EC.1. For $k_{T,T+t} \geq 2$,

$$\begin{aligned}
&\mathrm{DECT}(\Delta,t|r,\alpha,s,\beta,\gamma_{purch},\gamma_{attr},\mathbf{X}^P,\mathbf{X}^A,x,\mathbf{t},T)=\\
&= \frac{\prod_{j=1}^{x} A_{k_j}}{L(\alpha,r,\beta,s,\gamma_{purch},\gamma_{attr}|\mathbf{X}^P,\mathbf{X}^A,x,\mathbf{t},T)}\\
&\quad\times \frac{\Gamma(x+r)\alpha^r\beta^s}{\Gamma(r)\left(\sum_{j=1}^{x} B_{k_j}+B_{k_T}+\alpha\right)^{x+r}(D_{k_T}+\beta)^s}\\
&\quad\times \frac{\Delta^{s-1}(r+x)(D_{k_T}+\beta)^s}{\left(\sum_{j=1}^{x} B_{k_j}+B_{k_T}+\alpha\right)}\left[A_1/C_1^s S_1^*\right.\\
&\quad\left.+A_{k_{T,T+t}}/C^s_{k_{T,T+t}}S^*_{k_{T,T+t}}+\mathbb{I}_{\{3,4,\ldots\}}\times\sum_{i=2}^{k_{T,T+t}-1}A_i/C_i^s S_i\right],
\end{aligned} \tag{A.5}$$

with

$$\begin{aligned}
S_1^* &= \Psi\left(s,s;\frac{\Delta(C_1T+\bar{D}_1+\beta)}{C_1}\right)-\exp[-\Delta d_T]\\
&\quad\times\Psi\left(s,s;\frac{\Delta(C_1(T+d_T)+\bar{D}_1+\beta)}{C_1}\right),\\
S_i &= \exp[-\Delta(d_T+i-2)]\,\Psi\left(s,s;\frac{\Delta(C_ib_i^T+\bar{D}_i+\beta)}{C_i}\right)\\
&\quad-\exp[-\Delta(d_T+i-1)]\,\Psi\left(s,s;\frac{\Delta(C_i(b_i^T+1)+\bar{D}_i+\beta)}{C_i}\right),\\
&\quad i=2,\ldots,k_{T,T+t}-1\\
S^*_{k_{T,T+t}} &= \exp[-\Delta(d_T+k_{T,T+t}-2)]\\
&\quad\times\Psi\left(s,s;\frac{\Delta\left(C_{k_{T,T+t}}b^T_{k_{T,T+t}}+\bar{D}_{k_{T,T+t}}+\beta\right)}{C_{k_{T,T+t}}}\right)\\
&\quad-\exp[-\Delta(T+t-T)]\,\Psi\left(s,s;\frac{\Delta(C_{k_{T,T+t}}T+t+\bar{D}_{k_{T,T+t}}+\beta)}{C_{k_{T,T+t}}}\right),
\end{aligned}$$

$$
\begin{aligned}
D_{k_{T,t}} &= C_{k_{T,t}} t + \bar{D}_{k_{T,t}},\\
\bar{D}_i &= \exp(\gamma'_{attr}\mathbf{x}_1^A) d_2 + \sum_{l=2}^{k_{0,T}+i-2} \exp(\gamma'_{attr}\mathbf{x}_l^A)\\
&\quad + \exp\left(\gamma'_{attr}\mathbf{x}^A_{k_{0,T}+i-1}\right)\left[-d_2 - \delta(k_{0,T}+i-3)\right],\\
D_{k_T} &= D_{k_{T,0}} = C_{k_T} T + \bar{D}_{k_T},\\
A_i &= \exp\left(\gamma'_{purch}\mathbf{x}_i^P\right), \quad i = 1,\ldots,k_{T,T+t}\\
C_i &= \exp\left(\gamma'_{purch}\mathbf{x}^A_{k_{0,T}+i-1}\right),\\
b_i^T &= T + d_T + (i-2), \quad i = 2,..,k_{T,T+t}
\end{aligned}
$$

Ψ : is the confluent hypergeometric function of the second kind.

A.5. Unconditional Expectation $E[Y(t)]$

In the following, we present the closed-form solution for the unconditional expectation of the number of transactions in the interval $(0, t]$, $E[Y(t)]$, which is used for for the incremental tracking plots (Figures 2–4) to asses the aggregated performance of the model. We obtain the following number of transactions in $k_{0t} = 1$:

$$
\begin{aligned}
&E\left[Y(t)|r,\alpha,s,\beta,\gamma_{purch},\gamma_{attr},\mathbf{X}^P,\mathbf{X}^A\right]\\
&\quad = \frac{\beta^s r}{(s-1)\alpha}\left[\frac{A_1 t(s-1)}{(\beta + C_1 t)^s} + \frac{A_1/C_1}{\beta^{s-1}} - \frac{A_1(ts + 1/C_1\beta)}{(\beta + C_1 t)^s}\right],
\end{aligned}
$$

and for $k_{0t} \geq 2$,

$$
\begin{aligned}
&E\left[Y(t)|r,\alpha,s,\beta,\gamma_{purch},\gamma_{attr},\mathbf{X}^P,\mathbf{X}^A\right] =\\
&\quad = \frac{\beta^s r}{(s-1)\alpha}\left[\frac{B_{k_{0,t}}(s-1)}{(\beta + D_{k_{0,t}})^s} + S_1^* + S^*_{k_{0,t}} + \sum_{i=2}^{k_{0,t}-1} S_i\right],
\end{aligned}
$$

where

$$
\begin{aligned}
S_i &= \frac{A_i[b_i^0 s + 1/C_i(\beta + \bar{D}_i)] + \bar{B}_i(s-1)}{(\beta + \bar{D}_i + C_i b_i^0)^s}\\
&\quad - \frac{A_i[(b_i^0+1)s + 1/C_i(\beta + \bar{D}_i)] + \bar{B}_i(s-1)}{(\beta + \bar{D}_i + C_i(b_i^0+1))^s},\\
S_1^* &= \frac{A_1/C_1(\beta + \bar{D}_1) + \bar{B}_1(s-1)}{(\beta + \bar{D}_1)^s}\\
&\quad - \frac{A_1[d_1 s + 1/C_1(\beta + \bar{D}_1)] + \bar{B}_1(s-1)}{(\beta + \bar{D}_1 + C_1 d_1)^s},\\
S^*_{k_{0,t}} &= \frac{A_{k_{0,t}}\left[b^0_{k_{0,t}} s + 1/C_{k_{0,t}}(\beta + \bar{D}_{k_{0,t}})\right] + \bar{B}_{k_{0,t}}(s-1)}{\left(\beta + \bar{D}_{k_{0,t}} + C_{k_{0,t}} b^0_{k_{0,t}}\right)^s}\\
&\quad - \frac{A_{k_{0,t}}[ts + 1/C_{k_{0,t}}(\beta + \bar{D}_{k_{0,t}})] + \bar{B}_{k_{0,t}}(s-1)}{(\beta + \bar{D}_{k_{0,t}} + C_{k_{0,t}} t)^s},\\
B_{k_{0,t}} &= A_{k_{0,t}} t + \bar{B}_{k_{0,t}},\\
D_{k_{0,t}} &= C_{k_{0,t}} t + \bar{D}_{k_{0,t}},\\
\bar{B}_i &= \exp\left(\gamma'_{purch}\mathbf{X}_1^P\right) d_1 + \sum_{l=2}^{i-1} \exp\left(\gamma'_{purch}\mathbf{X}_l^P\right)\\
&\quad + \exp\left(\gamma'_{purch}\mathbf{X}_i^P\right)[-d_1 - \mathbb{I}\{i \geq 2\}(i-2)], \quad i = 1,\ldots,k_{0,t}\\
\bar{D}_i &= \exp(\gamma'_{attr}\mathbf{X}_1^A) d_2 + \sum_{l=2}^{i-1} \exp(\gamma'_{attr}\mathbf{X}_l^A) + \exp\\
&\quad \times (\gamma'_{attr}\mathbf{X}_i^A)[-d_2 - \mathbb{I}\{i \geq 2\}(i-2)], \quad i = 1,\ldots,k_{0,t}\\
b_i^0 &= d_1 + (i-2), \quad i = 2,..,k_{0,t}.
\end{aligned}
$$

A.6. Probability Mass Function $P[Y([u, t+u])) = x]$

In the following, we present the closed-form solution for the probability mass function. Additional derivations of the expression are provided in Online Appendix EC.2.

We obtain the following expression for the probability of the number of purchases in a given time interval $[u, u+t]$. The expression $P(Y((u, t+u]) = x|r, \alpha, s, \beta, \gamma_{purch}, \gamma_{attr}, \mathbf{X}^P, \mathbf{X}^A)$ is used for the in-sample histograms in Figures 2–4.

$$
\begin{aligned}
P(X((u,t+u]) &= x|r,\alpha,s,\beta,\gamma_{purch},\gamma_{attr},\mathbf{X}^P,\mathbf{X}^A)\\
&= S^1 + \frac{\Gamma(s+1)\Gamma(r+x)}{\Gamma(r+s+x+1)}\sum_{j=0}^{x}\left(S^2_{1,j} + S^2_{k_{u,t+u},j} + \sum_{i=2}^{k_{u,t+u}-1} S^2_{i,j}\right),
\end{aligned} \tag{A.6}
$$

with

$$
\begin{aligned}
S^1 &= \frac{B^x_{k_{u,t+u}}}{x!}\frac{\Gamma(x+r)}{\Gamma(r)}\frac{\alpha^r\beta^s}{(B_{k_{u,t+u}} + \alpha)^{x+r}(D_{k_{u,t+u}} + \beta)^s},\\
S^2_{i,j} &= \mathbb{I}\{\bar{B}_i + \alpha > (\bar{D}_i + \beta)A_i/C_i\} \times \bar{B}_i^j \frac{1}{j!}\sum_{n=0}^{x-j}\frac{1}{n!}\\
&\quad \times \frac{A_i^{s+n}}{C_i^s}\frac{\Gamma(r+s+j+n)}{\Gamma(r)\Gamma(s)}\left[\frac{[b_i^u]^n\alpha^r\beta^s}{(\bar{B}_i + A_i b_i^u + \alpha)^{r+s+j+n}}\right.\\
&\quad \times {}_2F_1\left(s+r+j+n, s+1; r+s+x+1;\right.\\
&\quad \left.\frac{\bar{B}_i + \alpha - (\bar{D}_i + \beta)A_i/C_i}{\bar{B}_i + A_i b_i^u + \alpha}\right) - \frac{[b_i^u + 1]^n\alpha^r\beta^s}{(\bar{B}_i + A_i(b_i^u+1) + \alpha)^{r+s+j+n}}\\
&\quad \times {}_2F_1\left(s+r+j+n, s+1; r+s+x+1;\right.\\
&\quad \left.\left.\frac{\bar{B}_i + \alpha - (\bar{D}_i + \beta)A_i/C_i}{\bar{B}_i + A_i(b_i^u+1) + \alpha}\right)\right]\\
&\quad + \mathbb{I}\{\bar{B}_i + \alpha < (\bar{D}_i + \beta)A_i/C_i\} \times \bar{B}_i^j \frac{1}{j!}\sum_{n=0}^{x-j}\frac{1}{n!}\\
&\quad \times \frac{\Gamma(r+s+j+n)}{\Gamma(r)\Gamma(s)}\frac{A_i^{s+n}}{C_i^s}\left[\frac{[b_i^u]^n\alpha^r\beta^s}{((\bar{D}_i + C_i b_i^u + \beta)A_i/C_i)^{r+s+j+n}}\right.\\
&\quad 2F_1\left(s+r+j+n, r+x; r+s+x+1;\right.\\
&\quad \left.\frac{(\bar{D}_i + \beta)A_i/C_i - (\bar{B}_i + \alpha)}{(\bar{D}_i + C_i b_i^u + \beta)A_i/C_i}\right)\\
&\quad - \frac{[b_i^u + 1]^n\alpha^r\beta^s}{((\bar{D}_i + C_i(b_i^u+1) + \beta)A_i/C_i)^{r+s+j+n}}\\
&\quad {}_2F_1\left(s+r+j+n, r+x; r+s+x+1;\right.\\
&\quad \left.\left.\frac{(\bar{D}_i + \beta)A_i/C_i - (\bar{B}_i + \alpha)}{(\bar{D}_i + C_i(b_i^u+1) + \beta)A_i/C_i}\right)\right]
\end{aligned}
$$

$$
\begin{aligned}
S^2_{1,j} &= \mathbb{I}\{\bar{B}_1+\alpha>(\bar{D}_1+\beta)A_1/C_1\}\\
&\quad\times\bar{B}_1^j\frac{1}{j!}\sum_{n=0}^{x-j}\frac{1}{n!}\frac{A_1^{s+n}}{C_1^s}\frac{\Gamma(r+s+j+n)}{\Gamma(r)\Gamma(s)}\Bigg[\frac{u^n\alpha^r\beta^s}{(\bar{B}_1+A_1u+\alpha)^{r+s+j+n}}\\
&\quad{}_2F_1\Bigg(s+r+j+n,s+1;r+s+x+1;\frac{\bar{B}_1+\alpha-(\bar{D}_1+\beta)A_1/C_1}{\bar{B}_1+A_1u+\alpha}\Bigg)-\frac{[b_2^u]^n\alpha^r\beta^s}{(\bar{B}_1+A_1b_2^u+\alpha)^{r+s+j+n}}\\
&\quad{}_2F_1\Bigg(s+r+j+n,s+1;r+s+x+1;\frac{\bar{B}_1+\alpha-(\bar{D}_1+\beta)A_1/C_1}{\bar{B}_1+A_1b_2^u+\alpha}\Bigg)\Bigg]\\
&\quad+\mathbb{I}\{\bar{B}_1+\alpha<(\bar{D}_1+\beta)A_1/C_1\}\\
&\quad\times\bar{B}_1^j\frac{1}{j!}\sum_{n=0}^{x-j}\frac{1}{n!}\frac{\Gamma(r+s+j+n)}{\Gamma(r)\Gamma(s)}\frac{A_1^{s+n}}{C_1^s}\\
&\quad\Bigg[\frac{u^n\alpha^r\beta^s}{((\bar{D}_1+C_1u+\beta)A_1/C_1)^{r+s+j+n}}\\
&\quad{}_2F_1\Bigg(s+r+j+n,r+x;r+s+x+1;\frac{(\bar{D}_1+\beta)A_1/C_1-(\bar{B}_1+\alpha)}{(\bar{D}_1+C_1u+\beta)A_1/C_1}\Bigg)\\
&\quad-\frac{[b_2^u]^n\alpha^r\beta^s}{((\bar{D}_1+C_1b_2^u+\beta)A_1/C_1)^{r+s+j+n}}\\
&\quad{}_2F_1\Bigg(s+r+j+n,r+x;r+s+x+1;\frac{(\bar{D}_1+\beta)A_1/C_1-(\bar{B}_1+\alpha)}{(\bar{D}_1+C_1b_2^u+\beta)A_1/C_1}\Bigg),
\end{aligned}
$$

$$
\begin{aligned}
S^2_{k_{u,t+u},j} &= \mathbb{I}\{\bar{B}_{k_{u,t+u}}+\alpha>(\bar{D}_{k_{u,t+u}}+\beta)A_{k_{u,t+u}}/C_{k_{u,t+u}}\}\\
&\quad\times\bar{B}^j_{k_{u,t+u}}\frac{1}{j!}\sum_{n=0}^{x-j}\frac{1}{n!}\frac{A^{s+n}_{k_{u,t+u}}}{C^s_{k_{u,t+u}}}\frac{\Gamma(r+s+j+n)}{\Gamma(r)\Gamma(s)}\\
&\quad\Bigg[\frac{\big[b^u_{k_{u,t+u}}\big]^n\alpha^r\beta^s}{\big(\bar{B}_{k_{u,t+u}}+A_{k_{u,t+u}}b^u_{k_{u,t+u}}+\alpha\big)^{r+s+j+n}}\\
&\quad{}_2F_1\Bigg(s+r+j+n,s+1;r+s+x+1;\frac{\bar{B}_{k_{u,t+u}}+\alpha-(\bar{D}_{k_{u,t+u}}+\beta)A_{k_{u,t+u}}/C_{k_{u,t+u}}}{\bar{B}_{k_{u,t+u}}+A_{k_{u,t+u}}b^u_{k_{u,t+u}}+\alpha}\Bigg)\\
&\quad-\frac{[u+t]^n\alpha^r\beta^s}{(\bar{B}_{k_{u,t+u}}+A_{k_{u,t+u}}(u+t)+\alpha)^{r+s+j+n}}\\
&\quad{}_2F_1\Bigg(s+r+j+n,s+1;r+s+x+1;\frac{\bar{B}_{k_{u,t+u}}+\alpha-(\bar{D}_{k_{u,t+u}}+\beta)A_{k_{u,t+u}}/C_{k_{u,t+u}}}{\bar{B}_{k_{u,t+u}}+A_{k_{u,t+u}}(u+t)+\alpha}\Bigg)\Bigg]\\
&\quad+\mathbb{I}\{\bar{B}_{k_{u,t+u}}+\alpha<(\bar{D}_{k_{u,t+u}}+\beta)A_{k_{u,t+u}}/C_{k_{u,t+u}}\}\\
&\quad\times\bar{B}^j_{k_{u,t+u}}\frac{1}{j!}\sum_{n=0}^{x-j}\frac{1}{n!}\frac{\Gamma(r+s+j+n)}{\Gamma(r)\Gamma(s)}\frac{A^{s+n}_{k_{u,t+u}}}{C^s_{k_{u,t+u}}}\\
&\quad\Bigg[\frac{\big[b^u_{k_{u,t+u}}\big]^n\alpha^r\beta^s}{\Big(\big(\bar{D}_{k_{u,t+u}}+C_{k_{u,t+u}}b^u_{k_{u,t+u}}+\beta\big)A_{k_{u,t+u}}/C_{k_{u,t+u}}\Big)^{r+s+j+n}}\\
&\quad{}_2F_1\Bigg(s+r+j+n,r+x;r+s+x+1;\frac{(\bar{D}_{k_{u,t+u}}+\beta)A_{k_{u,t+u}}/C_{k_{u,t+u}}-(\bar{B}_{k_{u,t+u}}+\alpha)}{\big(\bar{D}_{k_{u,t+u}}+C_{k_{u,t+u}}b^u_{k_{u,t+u}}+\beta\big)A_{k_{u,t+u}}/C_{k_{u,t+u}}}\Bigg)\\
&\quad-\frac{[u+t]^n\alpha^r\beta^s}{((\bar{D}_{k_{u,t+u}}+C_{k_{u,t+u}}(u+t)+\beta)A_{k_{u,t+u}}/C_{k_{u,t+u}})^{r+s+j+n}}\\
&\quad{}_2F_1\Bigg(s+r+j+n,r+x;r+s+x+1;\frac{(\bar{D}_{k_{u,t+u}}+\beta)A_{k_{u,t+u}}/C_{k_{u,t+u}}-(\bar{B}_{k_{u,t+u}}+\alpha)}{(\bar{D}_{k_{u,t+u}}+C_{k_{u,t+u}}(u+t)+\beta)A_{k_{u,t+u}}/C_{k_{u,t+u}}}\Bigg)\Bigg],
\end{aligned}
$$

for $i=1,\ldots,k_{u,t+u}$,

$$
\begin{aligned}
A_i &= \exp\big(\gamma'_{purch}\mathbf{X}_i^P\big),\\
C_i &= \exp\big(\gamma'_{attr}\mathbf{X}^A_{k_{0,u}+i-1}\big),\\
B_{k_{u,t+u}} &= A_{k_{u,t+u}}(t+u)+\bar{B}_{k_{u,t+u}},\\
D_{k_{u,t+u}} &= C_{k_{u,t+u}}t+\bar{D}_{u,t+u},\\
\bar{B}_i &= \exp\big(\gamma'_{purch}\mathbf{X}_1^P\big)d_1+\sum_{l=2}^{i-1}\exp\big(\gamma'_{purch}\mathbf{X}_l^P\big)\\
&\quad+\exp\big(\gamma'_{purch}\mathbf{X}_i^P\big)[-u-d_1-\mathbb{I}\{i\geq 2\}(i-2)],\\
\bar{D}_i &= \exp(\gamma'_{attr}\mathbf{X}_1^A)d_2+\sum_{l=2}^{k_{0,u}+i-2}\exp(\gamma'_{attr}\mathbf{X}_l^A)\\
&\quad+\exp\big(\gamma'_{attr}\mathbf{X}^A_{k_{0,u}+i-1}\big)\big[-d_2-\mathbb{I}\{k_{0,u}+i\geq 3\}(k_{0,u}+i-3)\big],
\end{aligned}
$$

and for $i=2,\ldots,k_{u,t+u}$,

$$b_i^u=u+d_1+i-2.$$

Appendix B. Additional Derivations for the Likelihood Expression

In this section, we show the detailed stop for deriving the closed-form likelihood expression. We start by distinguishing two cases that affect the purchase process:

Case 1. The customer is still alive at the end of the estimation period T. Let $\mathbf{t}=(t_1,\ldots,t_x)'$ denote the vector of transaction times and $z_{x,T}$ be the time between the last transaction of a customer and the end of the estimation period. The probability that the lifetime of a customer is larger than T, that is, with $Z_{x,x+1}$ as a random variable for the time between x and $x+1$, is

$$P(Z_{x,x+1}>z_{x,T}|\mathbf{t})=S^P(z_{x,T}|\mathbf{t})=\exp\big[-\theta^P_{t_x}(z_{x,T})\big]. \tag{B.1}$$

Then, the individual likelihood is the product of the density for the interpurchase time and the corresponding survivor function is

$$\begin{aligned}
L&\left(\lambda_0, \gamma_{purch}|\mathbf{t}, T, \mathbf{X}^P, x, \Omega > T\right)\\
&= \left[f^P(z_{0,1}) \cdot \prod_{j=2}^{x} f^P\left(z_{j-1,j}|z_{0,1}, \ldots, z_{j-2,j-1}\right)\right] S^P(z_{x,T}|\mathbf{t})\\
&= \left[\prod_{j=1}^{x} \lambda_0 A_{k_j} \exp\left[-\lambda_0 B_{k_j}\right]\right] \exp[-\lambda_0 B_{k_T}], \qquad \text{(B.2)}
\end{aligned}$$

with

$$\begin{aligned}
A_{k_j} &= \exp\left(\gamma'_{purch}\mathbf{x}^P_{k_j}\right),\\
B_{k_j} &= \exp\left(\gamma'_{purch}\mathbf{x}^P_1\right)d_1 + \left[\sum_{l=1}^{k_j-1} \exp\left(\gamma'_{purch}\mathbf{x}^P_l\right)\right]\\
&\quad + \exp\left(\gamma'_{purch}\mathbf{x}^P_{k_j}\right)\left[z_{j-1,j} - d_1 - \delta\left(k_j - 2\right)\right].
\end{aligned}$$

With k_T as the number of intervals between the last transaction and the end of the estimation period plus 1 and with the discretized version of $\theta_{t_x}(z_{x,T})$,

$$\begin{aligned}
B_{k_T} &= \exp\left(\gamma'_{purch}\mathbf{x}^P_1\right)d_1 + \left[\sum_{l=1}^{k_j-1} \exp\left(\gamma'_{purch}\mathbf{x}^P_l\right)\right]\\
&\quad + \exp\left(\gamma'_{purch}\mathbf{x}^P_{k_T}\right)[(T - t_x) - d_1 - \delta(k_T - 2)].
\end{aligned}$$

Case 2. The customer becomes inactive between the last transaction and T. This means that the customer's lifetime ends somewhere before the end of the estimation period (i.e., $\omega \in (t_x, T]$). Therefore, we need the probability for some fixed $\omega \in (t_x, T]$:

$$P(Z_{x,x+1} > z_{x,\omega}|\mathbf{t}) = S^P(z_{x,\omega}|\mathbf{t}) = \exp\left[-\theta^P_{t_x}(z_{x,\omega})\right]. \qquad \text{(B.3)}$$

Thus, we have the following likelihood for a single customer:

$$\begin{aligned}
L&\left(\lambda_0, \gamma_{purch}|\mathbf{t}, T, \mathbf{X}^P, x, \Omega \in (t_x, T]\right)\\
&= \left[f^P(z_{0,1}) \cdot \prod_{j=2}^{x} f^P\left(z_{j-1,j}|z_{0,1}, \ldots, z_{j-2,j-1}\right)\right] S^P(z_{x,\omega}|\mathbf{t})\\
&= \left[\prod_{j=1}^{x} \lambda_0 A_{k_j} \exp\left[-\lambda_0 B_{k_j}\right]\right] \exp[-\lambda_0 B_{k_{x,\omega}}], \qquad \text{(B.4)}
\end{aligned}$$

with A_{k_j} and B_{k_j} defined previously and the special case for $B_{k_{x,\omega}}$ where

$$\begin{aligned}
A_{k_{x,\omega}} &= \exp\left(\gamma'_{purch}\mathbf{x}^P_{k_{x,\omega}}\right),\\
\bar{B}_{k_{x,\omega}} &= \exp\left(\gamma'_{purch}\mathbf{x}^P_1\right)d_1 + \sum_{l=2}^{k_{x,\omega}-1} \exp\left(\gamma'_{purch}\mathbf{x}^P_l\right),\\
&\quad + \exp\left(\gamma'_{purch}\mathbf{x}^P_{k_{x,\omega}}\right)\left[-t_x - d_1 - \mathbb{I}\{k_{x,\omega} \geq 2\}(k_{x,\omega} - 2)\right]\\
B_{k_{x,\omega}} &= A_{k_{x,\omega}}\omega + \bar{B}_{k_{x,\omega}},
\end{aligned}$$

with $k_{x,\omega} = 1, \ldots, k_T$. For $\omega = T, B_{k_{x,\omega}} = B_{k_T}$. $\mathbb{I}\{k_{x,\omega} \geq 2\} = 1$, if $k_{x,\omega} \geq 2$, else 0. To combine the two cases, we multiply Case 1 by $P(\Omega > T)$ and take the integral from t_x to T of Case 2 multiplied by $f^A(\omega|\mu_0)$. Adding everything, we get Equation (19).

Finally, we introduce customer heterogeneity in (19) and allow the purchase and attrition rate to vary across customer (i.e., we allow λ_0 and μ_0 to vary across customers). Following Schmittlein et al. (1987), we assume the purchase rate Λ_0 and the attrition rate M_0 to be Gamma distributed with shape parameter r and scale parameter α, respectively, s and β,

$$g(\lambda_0) = \frac{\alpha^r \lambda_0^{r-1} e^{-\lambda_0 \alpha}}{\Gamma(r)} \text{ and } g(\mu_0) = \frac{\beta^s \mu_0^{s-1} e^{-\mu_0 \beta}}{\Gamma(s)}.$$

Thus,

$$\begin{aligned}
L&\left(\alpha, r, \beta, s, \gamma_{purch}, \gamma_{attr}|\mathbf{X}^P, \mathbf{X}^A, x, \mathbf{t}, T\right)\\
&= \int_0^\infty \int_0^\infty \left(\left[\prod_{j=1}^{x} \lambda_0 A_{k_j} \exp\left[-\lambda_0 B_{k_j}\right]\right]\right.\\
&\quad \times \exp[-\lambda_0 B_{k_T}] \exp[-\mu_0 D_{k_T}]\\
&\quad + \left[\prod_{j=1}^{x} \lambda_0 A_{k_j} \exp\left[-\lambda_0 B_{k_j}\right]\mu_0\right]\\
&\quad \left.\times \left[I_1^*(\lambda_0, \mu_0) + I_{k_T}^*(\lambda_0, \mu_0) + \sum_{i=2}^{k_T-1} I_i(\lambda_0, \mu_0)\right]\right)\\
&\quad \times g(\lambda_0)g(\mu_0)d\lambda_0 d\mu_0. \qquad \text{(B.5)}
\end{aligned}$$

By removing the conditioning on λ_0 and μ_0, we are able to derive the final individual likelihood:

$$\begin{aligned}
L&\left(\alpha, r, \beta, s, \gamma_{purch}, \gamma_{attr}|\mathbf{X}^P, \mathbf{X}^A, x, \mathbf{t}, T\right)\\
&= \int_0^\infty \int_0^\infty \left(\left[\prod_{j=1}^{x} \lambda_0 A_{k_j} \exp\left[-\lambda_0 B_{k_j}\right]\right]\right.\\
&\quad \times \exp[-\lambda_0 B_{k_T}] \exp[-\mu_0 D_{k_T}]\\
&\quad + \left[\prod_{j=1}^{x} \lambda_0 A_{k_j} \exp\left[-\lambda_0 B_{k_j}\right]\mu_0\right]\\
&\quad \left.\times \left[I_1^*(\lambda_0, \mu_0) + I_{k_T}^*(\lambda_0, \mu_0) + \sum_{i=2}^{k_T-1} I_i(\lambda_0, \mu_0)\right]\right)\\
&\quad \times g(\lambda_0)g(\mu_0)d\lambda_0 d\mu_0. \qquad \text{(B.6)}
\end{aligned}$$

Note that (B.6) equates (B.5) presented in the main body of the paper. We solve the integral separately for the two summands. For the first summand, we obtain

$$\begin{aligned}
&\int_0^\infty \int_0^\infty \left[\prod_{j=1}^{x} \lambda_0 A_{k_j} \exp\left[-\lambda_0 B_{k_j}\right]\right] \exp[-\lambda_0 B_{k_T}] \exp[-\mu_0 D_{k_T}]\\
&\qquad \times g(\lambda_0)g(\mu_0)d\lambda_0 d\mu_0\\
&= \prod_{j=1}^{x} A_{k_j} \frac{\Gamma(x + r)}{\Gamma(r)} \frac{\alpha^r \beta^s}{\left(\sum_{j=1}^{x} B_{k_j} + B_{k_T} + \alpha\right)^{x+r} \left(D_{k_T} + \beta\right)^s}.
\end{aligned}$$

(B.7)

For the second summand, we obtain

$$\begin{aligned}&\int_0^\infty\int_0^\infty\left[\prod_{j=1}^{x}\lambda_0 A_{k_j}\exp\left[-\lambda_0 B_{k_j}\right]\mu_0\right]\\&\quad\times\left[I_1^*(\lambda_0,\mu_0)+I_{k_T}^*(\lambda_0,\mu_0)+\sum_{i=2}^{k_T-1}I_i(\lambda_0,\mu_0)\right]\\&\quad\times g(\lambda_0)g(\mu)d\lambda_0 d\mu_0\\&=\prod_{j=1}^{x}A_{k_j}\int_0^\infty\int_0^\infty\lambda_0^x\exp\left[-\lambda_0\sum_{j=1}^{x}B_{k_j}\right]\\&\quad\times\left[\mu_0 I_1^*(\lambda_0,\mu_0)+\mu_0 I_{k_T}^*(\lambda_0,\mu_0)+\sum_{i=2}^{k_T-1}\mu_0 I_i(\lambda_0,\mu_0)\right]\\&\quad\times g(\lambda_0)g(\mu_0)d\lambda_0 d\mu_0.\end{aligned}\tag{B.8}$$

For one I_i, $i=2,\dots,k_T-1$, we obtain

$$\begin{aligned}&\int_0^\infty\int_0^\infty\lambda_0^x\exp\left[-\lambda_0\sum_{j=1}^{x}B_{k_j}\right]\mu_0 I_i(\lambda_0,\mu)g(\lambda_0)g(\mu)d\lambda_0 d\mu_0\\&=C_i\int_0^\infty\int_0^\infty\frac{\lambda_0^x\mu_0}{\lambda_0 A_i+\mu_0 C_i}\\&\quad\times\left(\exp\left[-\left(\lambda_0\left(\sum_{j=1}^{x}B_{k_j}+\bar{B}_i+(t_x+d_1+(i-2))A_i\right)\right.\right.\right.\\&\quad\left.\left.+\mu_0\left(\bar{D}_i+(t_x+d_1+(i-2))C_i\right)\right)\right]\\&\quad-\exp\left[-\left(\lambda_0\left(\sum_{j=1}^{x}B_{k_j}+\bar{B}_i+(t_x+d_1+(i-1))A_i\right)\right.\right.\\&\quad\left.\left.\left.+\mu_0\left(\bar{B}_i+(t_x+d_1+(i-1))C_i\right)\right)\right]\right)\\&\quad\times g(\lambda_0)g(\mu_0)d\lambda_0 d\mu_0.\end{aligned}\tag{B.9}$$

Using the integral expression form in Online Appendix EC.3, for $a_i+\alpha\geq(b_i+\beta)A_i/C_i$, we obtain

$$\begin{aligned}&=\frac{\alpha^r\beta^s}{r+s+x}\frac{\Gamma(s+1)\Gamma(r+x)}{\Gamma(r)\Gamma(s)}\frac{A_i^s}{C_i^s}\left[\frac{1}{(a_i+\alpha)^{r+s+x}}\right.\\&\quad\times{}_2F_1\left(r+s+x,s+1;r+s+x+1;\right.\\&\quad\left.\frac{a_i+\alpha-(b_i+\beta)A_i/C_i}{a_i+\alpha}\right)-\frac{1}{((a_i+A_i)+\alpha)^{r+s+x}}\\&\quad{}_2F_1\left(r+s+x,s+1;r+s+x+1;\right.\\&\quad\left.\left.\frac{(a_i+A_i)+\alpha-((b_i+C_i)+\beta)A_i/C_i}{(a_i+A_i)+\alpha}\right)\right],\end{aligned}\tag{B.10}$$

and if $a_i+\alpha\leq(b_i+\beta)A_i/C_i$, then

$$\begin{aligned}&=\frac{\alpha^r\beta^s}{r+s+x}\frac{\Gamma(s+1)\Gamma(r+x)}{\Gamma(r)\Gamma(s)}\frac{A_i^s}{C_i^s}\left[\frac{1}{((b_i+\beta)A_i/C_i)^{r+s+x}}\right.\\&\quad{}_2F_1\left(r+s+x,r+x;r+s+x+1;\frac{(b_i+\beta)A_i/C_i-(a_i+\alpha)}{(b_i+\beta)A_i/C_i}\right)\\&\quad-\frac{1}{(((b_i+C_i)+\beta)A_i/C_i)^{r+s+x}}\\&\quad{}_2F_1\left(r+s+x,r+x;r+s+x+1;\right.\\&\quad\left.\left.\frac{((b_i+C_i)+\beta)A_i/C_i-((a_i+A_i)+\alpha)}{((b_i+C_i)+\beta)A_i/C_i}\right)\right],\end{aligned}\tag{B.11}$$

with

$$a_i=\sum_{j=1}^{x}B_{k_j}+\bar{B}_i+(t_x+d_1+(i-2))A_i,\qquad i=1,\dots,k_T$$
$$b_i=\bar{D}_i+(t_x+d_1+(i-2))C_i,\qquad i=1,\dots,k_T.$$

Similarly for $I_1^*(\lambda_0,\mu_0)$, we get for $(a_1+A_1)+\alpha\geq((b_1+C_1)+\beta)A_1/C_1$,

$$\begin{aligned}&=\frac{\alpha^r\beta^s}{r+s+x}\frac{\Gamma(s+1)\Gamma(r+x)}{\Gamma(r)\Gamma(s)}\frac{A_1^s}{C_1^s}\left[\frac{1}{((a_1+(1-d_1)A_1)+\alpha)^{r+s+x}}\right.\\&\quad{}_2F_1\left(r+s+x,s+1;r+s+x+1;\right.\\&\quad\left.\frac{(a_1+(1-d_1)A_1)+\alpha-((b_1+(1-d_1)C_1)+\beta)A_1/C_1}{(a_1+(1-d_1)A_1)+\alpha}\right)\\&\quad-\frac{1}{((a_1+A_1)+\alpha)^{r+s+x}}\\&\quad{}_2F_1\left(r+s+x,s+1;r+s+x+1;\right.\\&\quad\left.\left.\frac{(a_1+A_1)+\alpha-((b_1+C_1)+\beta)A_1/C_1}{(a_1+A_1)+\alpha}\right)\right],\end{aligned}\tag{B.12}$$

whereas for $(a_1+A_1)+\alpha\leq((b_1+C_1)+\beta)A_1/C_1$,

$$\begin{aligned}&=\frac{\alpha^r\beta^s}{r+s+x}\frac{\Gamma(s+1)\Gamma(r+x)}{\Gamma(r)\Gamma(s)}\frac{A_1^s}{C_1^s}\left[\frac{1}{((b_1+(1-d_1)C_1)+\beta)A_1/C_1)^{r+s+x}}\right.\\&\quad{}_2F_1\left(r+s+x,r+x;r+s+x+1;\right.\\&\quad\left.\frac{((b_1+(1-d_1)C_1)+\beta)A_1/C_1-((a_1+(1-d_1)A_1)+\alpha)}{((b_1+(1-d_1)C_1)+\beta)A_1/C_1}\right)\\&\quad-\frac{1}{(((b_1+C_1)+\beta)A_1/C_1)^{r+s+x}}\\&\quad{}_2F_1\left(r+s+x,r+x;r+s+x+1;\right.\\&\quad\left.\left.\frac{((b_1+C_1)+\beta)A_1/C_1-((a_1+A_1)+\alpha)}{((b_1+C_1)+\beta)A_1/C_1}\right)\right],\end{aligned}\tag{B.13}$$

where a_1, b_1 are defined previously. In the same manner for $I^*_{k_T}(\lambda_0, \mu)$ with $a_{k_T} + \alpha \geq (b_{k_T} + \beta)A_{k_T}/C_{k_T}$,

$$= \frac{\alpha^r \beta^s}{r+s+x} \frac{\Gamma(s+1)\Gamma(r+x)}{\Gamma(r)\Gamma(s)} \frac{A^s_{k_T}}{C^s_{k_T}} \Bigg[\frac{1}{(a_{k_T}+\alpha)^{r+s+x}} \,{}_2F_1\left(r+s+x, s+1; r+s+x+1; \frac{a_{k_T}+\alpha-(b_{k_T}+\beta)A_{k_T}/C_{k_T}}{a_{k_T}+\alpha}\right) - \frac{1}{(a^*_T+\alpha)^{r+s+x}} \,{}_2F_1\left(r+s+x, s+1; r+s+x+1; \frac{a^*_T+\alpha-(b^*_T+\beta)A_{k_T}/C_{k_T}}{a^*_T+\alpha}\right)\Bigg], \tag{B.14}$$

whereas for $a_{k_T} + \alpha \leq (b_{k_T} + \beta)A_{k_T}/C_{k_T}$,

$$= \frac{\alpha^r \beta^s}{r+s+x} \frac{\Gamma(s+1)\Gamma(r+x)}{\Gamma(r)\Gamma(s)} \frac{A^s_{k_T}}{C^s_{k_T}} \Bigg[\frac{1}{((b_{k_T}+\beta)A_{k_T}/C_{k_T})^{r+s+x}} \,{}_2F_1\left(r+s+x, r+x; r+s+x+1; \frac{(b_{k_T}+\beta)A_{k_T}/C_{k_T}-(a_{k_T}+\alpha)}{(b_{k_T}+\beta)A_{k_T}/C_{k_T}}\right) - \frac{1}{((b^*_T+\beta)A_{k_T}/C_{k_T})^{r+s+x}} \,{}_2F_1\left(r+s+x, r+x; r+s+x+1; \frac{(b^*_T+\beta)A_{k_T}/C_{k_T}-(a^*_T+\alpha)}{(b^*_T+\beta)A_{k_T}/C_{k_T}}\right)\Bigg], \tag{B.15}$$

with

$$a^*_T = \sum_{j=1}^{x} B_{k_j} + \bar{B}_{k_T} + TA_{k_T},$$
$$b^*_T = \bar{D}_{k_T} + TC_{k_T},$$

respectively. The term ${}_2F_1(a, b; c; z)$ is the Gaussian hypergeometric function of the second kind. Combining these parts results in the closed-form expression reported in Appendix A.1. In Online Appendix EC.4, we show that the extended Pareto/NBD model nests the standard Pareto/NBD model.

Appendix C. Additional Derivations for the Conditional Expectation

In this section, we present additional derivations for the expression of the conditional expectation. In particular, step-by-step solutions to two integral expressions shown in the main body of the paper are presented. This section is based on the Section 3.5, where the concept of the conditional expectation is introduced. See Section 3.5, Table A.1, and Appendix A.1 for specifications.

We start by deriving the conditional expectation under the assumption that the customer is alive after the end of the estimation period T. As such, we define

$$\lambda_l = \lambda_0 \exp\left(\gamma'_{purch} \mathbf{x}^P_l\right),$$
$$A_{k_{T,\omega}} = \exp\left(\gamma'_{purch} \mathbf{x}^P_{k_{T,\omega}}\right),$$
$$\bar{B}_{k_{T,\omega}} = \exp\left(\gamma'_{purch} \mathbf{x}^P_1\right) d_1 + \sum_{l=2}^{k_{T,\omega}-1} \exp\left(\gamma'_{purch} \mathbf{x}^P_l\right) + \exp\left(\gamma'_{purch} \mathbf{x}^P_{k_{T,\omega}}\right)[-T - d_1 - \delta(k_{T,\omega} - 2)],$$
$$B_{k_{T,\omega}} = A_{k_{T,\omega}} \omega + \bar{B}_{k_{T,\omega}},$$

and $k_{T,T+t}$ as the number of time intervals between T and $T + t$ including both of them.

Let $\{N(t)\}_{t \geq 0}$ be a nonhomogenous Poisson process with a bounded rate $\lambda(t)$. Then, for the number of events in the interval $(s, t+s]$ with $0 \leq s < t$, $N(t+s) - N(t)$,

$$N(t+s) - N(t) \sim \text{Poi}(m(t)),$$

where $m(t) = \int_s^{t+s} \lambda(y)\, dy$. Although we assume that all customers are alive at the end of the estimation period, there are still two different cases to consider about the timing of customer attrition.

Case 1. The lifetime of the customer is longer than the prediction period, that is, $\omega > T + t$. In this case, for $\omega > T + t$ and $l = 1, \ldots, k_{T,T+t}$, we get

$$E\left[Y(T, T+t) | \lambda_0, \mathbf{X}^P, \gamma_{purch}, \Omega > T+t\right] = \lambda_1 d_1 + \sum_{l=2}^{k_{T,T+t}-1} \lambda_l + \lambda_{k_{T,T+t}}[t_{T,T+t} - d_1 - \delta(k_{T,T+t} - 2)] = \lambda_0 B_{k_t, T+t}.$$

Case 2. The customer stops purchasing between the end of the observation period and the end of the prediction period, that is, $\omega \in (T, T+t]$. In this case, we get the following with $t_{T,\omega} = \omega - T$:

$$E\left[Y(T, T+t) | \lambda_0, \mathbf{X}^P, \gamma_{purch}, \omega \in (T, T+t]\right] = \lambda_1 d_1 + \sum_{l=2}^{k_{T,\omega}-1} \lambda_l + \lambda_{k_{T,\omega}}[t_{T,\omega} - d_1 - \delta(k_{T,\omega} - 2)] = \lambda_0 \exp\left(\gamma'_{purch} \mathbf{x}^P_{k_{T,\omega}}\right)\omega + \lambda_0 \bar{B}_{k_{T,\omega}}.$$

Combining the two cases that distinguish whether a customer is still active at the end of the prediction period (i.e., $\omega > T + t$) or has stopped purchasing between the end of the estimation period and the end of the prediction period (i.e., $\omega \in [T, T+t)$), we derive the following expression:

$$\begin{aligned} &E\left[Y(T, T+t) | \lambda_0, \mu_0, \mathbf{X}^P, \mathbf{X}^A, \gamma_{purch}, \gamma_{attr}, x, \mathbf{t}, \Omega > T\right] \\ &= E\left[Y(T, T+t) | \lambda_0, \mu_0, \mathbf{X}^P, \mathbf{X}^A, \gamma_{purch}, \gamma_{attr}, \Omega > T\right] \\ &= E\left[Y(T, T+t) | \lambda_0, \mathbf{X}^P, \gamma_{purch}, \Omega > T+t\right] \\ &\quad \times P(\Omega > t+T | \mu_0, \Omega > T) + \\ &\quad + \int_T^{T+t} \left(E\left[Y(T, T+t) | \lambda_0, \mathbf{X}^P, \gamma_{purch}, \omega \in (T, T+t]\right]\right) \\ &\quad \times f(\omega | \mu_0, \Omega > T)\, d\omega \\ &= \lambda_0 B_{k_{T,T+t}} P(\Omega > t+T | \mu_0, \Omega > T) + \\ &\quad + \int_T^{T+t} \left(\lambda_0 \exp\left(\gamma'_{purch} \mathbf{x}^P{}_{k_{T,\omega}}\right)\omega + \lambda_0 \bar{B}_{k_{T,\omega}}\right) \\ &\quad \times f(\omega | \mu_0, \Omega > T)\, d\omega. \end{aligned} \tag{C.1}$$

Note that (C.1) is the same as Equation (24) in the main body of the paper.

To solve (C.1), we define for the attrition process:

$$C_{k_{T,\omega}} = \exp\left(\gamma'_{attr} \mathbf{x}^A_{k_{0,T}+k_{T,\omega}-1}\right),$$
$$D_{k_{T,\omega}} = \exp\left(\gamma'_{attr} \mathbf{x}^A_{k_{0,T}+k_{T,\omega}-1}\right)\omega + \bar{D}_{k_{T,\omega}},$$

with

$$\bar{D}_{k_{T,\omega}} = \exp(\gamma'_{attr}\mathbf{x}_1^A)d_2 + \sum_{l=2}^{k_{0,T}+k_{T,\omega}-2} \exp(\gamma'_{attr}\mathbf{x}_l^A) + \exp\left(\gamma'_{attr}\mathbf{x}^A_{k_{0,T}+k_{T,\omega}-1}\right)[-d_2 - \delta(k_{0,T}+k_{T,\omega}-3)].$$

For the purchase process, we again use the *relative notation*, where the interval count starts for every purchase from one.

The conditional density in the integral in (C.1) is then given as

$$\begin{aligned} f^A(\omega|\mu_0,\Omega > T) &= \frac{f^A(\omega,\Omega > T|\mu_0)}{P(\Omega > T|\mu_0)} \\ &= \mu_0 C_{k_{T,\omega}} \exp[-\mu_0(C_{k_{T,\omega}}\omega + \bar{D}_{k_{T,\omega}} - D_{k_T})] \\ &\quad \times \mathbb{I}\{\omega > T\}, \end{aligned} \tag{C.2}$$

where $\mathbb{I}\{\omega > T\} = 1$, if $\omega \in (T,\infty)$ and 0 else. Similarly, the conditional probability can be derived as

$$\begin{aligned} P(\Omega > t+T|\mu_0,\Omega > T) &= \frac{P(\Omega > t+T,\Omega > T|\mu_0)}{P(\Omega > T|\mu_0)} \\ &= \exp[-\mu_0(D_{k_{T,T+t}} - D_{k_T})], \end{aligned}$$

where

$$\begin{aligned} D_{k_{T,T+t}} &= C_{k_{T,T+t}}(T+t) + \bar{D}_{k_{T,T+t}}, \\ D_{k_T} &= \exp\left(\gamma'_{attr}\mathbf{x}^A_{k_{0,x}+k_T-1}\right)T + \exp(\gamma'_{attr}\mathbf{x}_1^A) \\ &\quad \times d_2 \sum_{l=2}^{k_{0,x}+k_T-2} \exp(\gamma'_{attr}\mathbf{x}_l^A) + \\ &\quad + \exp\left(\gamma'_{attr}\mathbf{x}^A_{k_{0,x}+k_T-1}\right)[-d_2 - \mathbb{I}\{k_{0,x}.+k_T - 1 \geq 2\} \\ &\quad \times (k_{0,x}+k_T-3)]. \end{aligned}$$

We now focus on solving the integral in (C.1). Unfortunately, this is cumbersome, as we are forced to split the range of integration. In particular,

$$\begin{aligned} &\int_T^{T+t} (\lambda_0 A_{k_{T,\omega}}\omega + \lambda_0 \bar{B}_{k_{T,\omega}})\mu_0 C_{k_{T,\omega}} \\ &\quad \times \exp[-\mu_0(C_{k_{T,\omega}}\omega + \bar{D}_{k_{T,\omega}} - D_{k_T})]\, d\omega \\ &= \int_T^{T+d_1} \lambda_0(A_1\omega + \bar{B}_1)\mu_0 C_1 \exp[-\mu_0(C_1\omega + \bar{D}_1 - D_{k_T})]\, d\omega \\ &\quad + \int_{T+d_1+(k_{T,T+t}-2)}^{T+t} \lambda_0(A_{k_{T,T+t}}\omega + \bar{B}_{k_{T,T+t}})\mu_0 C_{k_{T,T+t}} \\ &\quad \exp[-\mu_0(C_{k_{T,T+t}}\omega + \bar{D}_{k_{T,T+t}} - D_{k_T})]\, d\omega \\ &\quad + \sum_{i=2}^{k_{T,T+t}-1} \int_{T+d_1+(i-2)}^{T+d_1+(i-1)} \lambda_0(A_i\omega + \bar{B}_i)\mu_0 C_i \\ &\quad \times \exp[-\mu_0(C_i\omega + \bar{D}_i - D_{k_T})]\, d\omega = \sum_{i=1}^{k_{T,T+t}} J_i^*(\lambda_0,\mu_0), \end{aligned} \tag{C.3}$$

with

$$\begin{aligned} J_i^*(\lambda_0,\mu_0) &= \lambda_0 A_i \Bigg(\left(b_i^T - (d_1-1)\mathbb{I}_{\{i=1\}}\right) e^{-\mu_0 C_i\left(b_i^T - (d_1-1)\mathbb{I}_{\{i=1\}}\right)} \\ &\quad - \left((b_i^T+1)\mathbb{I}_{\{i<k_{T,T+t}\}} + (T+t)\mathbb{I}_{\{i=k_{T,T+t}\}} \right) \\ &\quad \times e^{-\mu_0 C_i\left((b_i^T+1)\mathbb{I}_{\{i<k_{T,T+t}\}} + (T+t)\mathbb{I}_{\{i=k_{T,T+t}\}}\right)} \\ &\quad - \frac{1}{\mu_0 C_i}\Bigg[e^{-\mu_0 C_i\left((b_i^T+1)\mathbb{I}_{\{i<k_{T,T+t}\}} + (T+t)\mathbb{I}_{\{i=k_{T,T+t}\}}\right)} \\ &\quad - e^{-\mu_0 C_i\left(b_i^T-(d_1-1)\mathbb{I}_{\{i=1\}}\right)} \Bigg]\Bigg) + \lambda_0 \bar{B}_i \Bigg(e^{-\mu_0 C_i\left(b_i^T - (d_1-1)\mathbb{I}_{\{i=1\}}\right)} \\ &\quad - e^{-\mu_0 C_i\left((b_i^T+1)\mathbb{I}_{\{i<k_{T,T+t}\}} + (T+t)\mathbb{I}_{\{i=k_{T,T+t}\}}\right)} \Bigg), \end{aligned}$$

where

$$b_i^T = T + d_1 + (i-2)\ ,\ i = 2,..,k_{T,T+t}.$$

Thus, we finally arrive at

$$\begin{aligned} &E[Y(T,T+t)|\lambda_0,\mu_0,\mathbf{X}^P,\mathbf{X}^A,\gamma_{purch},\gamma_{attr},x,\mathbf{t},\Omega > T] \\ &\quad = \lambda_0 B_{k_{T,T+t}} \exp[-\mu_0(D_{k_{T,T+t}} - D_{k_T})] \\ &\quad + \sum_{i=1}^{k_{T,T+t}} e^{-\mu_0(\bar{D}_i - D_{k_T})} J_i^*(\lambda_0,\mu_0). \end{aligned} \tag{C.4}$$

To relax this assumption of customers being alive at the end of the estimation period T, we add the posterior distribution of being alive at T (see Equation (A.2)):

$$\begin{aligned} &E[Y(T,T+t)|\lambda_0,\mu_0,\mathbf{X}^P,\mathbf{X}^A,\gamma_{purch},\gamma_{attr},x,\mathbf{t},T] \\ &\quad = E[Y(T,T+t)|\lambda_0,\mu_0,\mathbf{X}^P,\mathbf{X}^A,\gamma_{purch},\gamma_{attr},x,\mathbf{t},\Omega > T] \\ &\quad \times P(\Omega > T|\lambda_0,\mu_0,\mathbf{X}^P,\mathbf{X}^A,\gamma_{purch},\gamma_{attr},x,\mathbf{t},T). \end{aligned} \tag{C.5}$$

Note (C.5) is the same as (25) in the main body of the paper. To arrive at the closed-form solution, we need to remove the conditioning on λ_0 and μ_0:

$$\begin{aligned} &E[Y(T,T+t)|r,\alpha,s,\beta,\gamma_{purch},\gamma_{attr},\mathbf{X}^P,\mathbf{X}^A,x,\mathbf{t},T] \\ &= \int_0^\infty \int_0^\infty E[Y(T,T+t)|\lambda_0,\mu_0,\mathbf{X}^P,\mathbf{X}^A,\gamma_{purch},\gamma_{attr},\Omega > T] \\ &\quad \times P(\Omega > T|\lambda_0,\mu_0,\mathbf{X}^P,\mathbf{X}^A,\gamma_{purch},\gamma_{attr},x,\mathbf{t},T) \\ &\quad \times g(\lambda_0,\mu_0|r,\alpha,s,\beta,\gamma_{purch},\gamma_{attr},\mathbf{X}^P,\mathbf{X}^A,x,\mathbf{t},T)\, d\lambda_0 d\mu_0 \\ &= \frac{\prod_{j=1}^x A_{k_j}}{L(\alpha,r,\beta,s,\gamma_{purch},\gamma_{attr}|\mathbf{X}^P,\mathbf{X}^A,x,\mathbf{t},T)} \frac{\alpha^r\beta^s}{\Gamma(r)\Gamma(s)} \\ &\quad \times \int_0^\infty \int_0^\infty \Bigg[\lambda_0 B_{k_{T,T+t}} \exp[-\mu_0(D_{k_{T,T+t}} - D_{k_T})] \\ &\quad + \sum_{i=1}^{k_{T,T+t}} e^{-\mu_0(\bar{D}_i - D_{k_T})} J_i^*(\lambda_0,\mu_0) \Bigg] \lambda_0^{x+r-1} \\ &\quad \times \exp\left[-\lambda_0\left(\sum_{j=1}^x B_{k_j} + B_{k_T} + \alpha\right)\right] \times \exp(-\mu_0(D_{k_T}+\beta)). \end{aligned} \tag{C.6}$$

To solve (C.6), we ignore the constant term for the moment and look at the summands of the integral separately. First, we solve

$$\int_0^\infty \int_0^\infty \lambda_0 B_{k_{T,T+t}} \exp[-\mu_0(D_{k_{T,T+t}} - D_{k_T})]\lambda_0^{x+r-1} \exp\left[-\lambda_0\left(\sum_{j=1}^{x} B_{k_j} + B_{k_T} + \alpha\right)\right] \exp(-\mu_0(D_{k_T} + \beta))\mu_0^{s-1}\, d\lambda_0 d\mu_0 = B_{k_{T,T+t}} \frac{\Gamma(x+r+1)}{\left(\sum_{j=1}^{x} B_{k_j} + B_{k_T} + \alpha\right)^{x+r+1}} \frac{\Gamma(s)}{(D_{k_{T,T+t}} + \beta)^s}. \tag{C.7}$$

We solve the second summand for $i = 1, \ldots, k_{T,T+t}$. However, we look at three cases separately: (1) $i = 2, \ldots, k_{T,T+t} - 1$, (2) $i = 1$, and (3) $i = k_{T,T+t}$.

Case 1. We first look at the middle part consisting of $i = 2, \ldots, k_{T,T+t} - 1$. The calculation is tedious, but straightforward, as $\frac{\Gamma(s)}{\Gamma(s-1)} = s - 1$:

$$\begin{aligned}&\int_0^\infty \int_0^\infty e^{-\mu_0(\bar{D}_i - D_{k_T})} J_i^*(\lambda_0, \mu_0)\lambda_0^{x+r-1} \exp\left[-\lambda_0\left(\sum_{j=1}^{x} B_{k_j} + B_{k_T} + \alpha\right)\right] \\ &\times \exp(-\mu_0(D_{k_T} + \beta))\mu_0^{s-1}\, d\lambda_0 d\mu_0 \\ &= \frac{\Gamma(x+r+1)\Gamma(s-1)}{\left(\sum_{j=1}^{x} B_{k_j} + B_{k_T} + \alpha\right)^{x+r+1}} \\ &\times \left[\frac{A_i[b_i^T s + 1/C_i(\bar{D}_i + \beta)] + \bar{B}_i(s-1)}{(\bar{D}_i + \beta + C_i b_i^T)^s} - \frac{A_i[(b_i^T + 1)s + 1/C_i(\bar{D}_i + \beta)] + \bar{B}_i(s-1)}{(\bar{D}_i + \beta + C_i(b_i^T + 1))^s}\right] \\ &:= \frac{\Gamma(x+r+1)\Gamma(s-1)}{\left(\sum_{j=1}^{x} B_{k_j} + B_{k_T} + \alpha\right)^{x+r+1}} S_i.\end{aligned}$$

Case 2. The first part $i = 1$:

$$\begin{aligned}&\int_0^\infty \int_0^\infty e^{-\mu_0(\bar{D}_1 - D_{k_T})} J_1^*(\lambda_0, \mu_0)\lambda_0^{x+r-1} \exp\left[-\lambda_0\left(\sum_{j=1}^{x} B_{k_j} + B_{k_T} + \alpha\right)\right] \exp(-\mu_0(D_{k_T} + \beta))\mu_0^{s-1}\, d\lambda_0 d\mu_0 \\ &= \frac{\Gamma(x+r+1)\Gamma(s-1)}{\left(\sum_{j=1}^{x} B_{k_j} + B_{k_T} + \alpha\right)^{x+r+1}} \\ &\times \left[\frac{A_1[Ts + 1/C_1(\bar{D}_1 + \beta)] + \bar{B}_1(s-1)}{(\bar{D}_1 + \beta + C_1 T)^s} - \frac{A_1[(T + d_1)s + 1/C_1(\bar{D}_1 + \beta)] + \bar{B}_1(s-1)}{(\bar{D}_1 + \beta + C_1(T + d_1))^s}\right] \\ &:= \frac{\Gamma(x+r+1)\Gamma(s-1)}{\left(\sum_{j=1}^{x} B_{k_j} + B_{k_T} + \alpha\right)^{x+r+1}} S_1^*.\end{aligned}$$

Case 3. Finally for $i = k_{T,T+t}$,

$$\begin{aligned}&\int_0^\infty \int_0^\infty e^{-\mu_0(\bar{D}_{k_{T,T+t}} - D_{k_T})} J_{k_{T,T+t}}^*(\lambda_0, \mu_0)\lambda_0^{x+r-1} \\ &\times \exp\left[-\lambda_0\left(\sum_{j=1}^{x} B_{k_j} + B_{k_T} + \alpha\right)\right] \exp(-\mu_0(D_{k_T} + \beta))\mu_0^{s-1}\, d\lambda_0 d\mu_0 \\ &= \frac{\Gamma(x+r+1)\Gamma(s-1)}{\left(\sum_{j=1}^{x} B_{k_j} + B_{k_T} + \alpha\right)^{x+r+1}} \\ &\times \left[\frac{A_{k_{T,T+t}}\left[b_{k_{T,T+t}}^T s + 1/C_i(\bar{D}_{k_{T,T+t}} + \beta)\right] + \bar{B}_{k_{T,T+t}}(s-1)}{\left(\bar{D}_{k_{T,T+t}} + \beta + C_{k_{T,T+t}} b_{k_{T,T+t}}^T\right)^s} - \frac{A_{k_{T,T+t}}[(T+t)s + 1/C_{k_{T,T+t}}(\bar{D}_{k_{T,T+t}} + \beta)] + \bar{B}_{k_{T,T+t}}(s-1)}{(\bar{D}_{k_{T,T+t}} + \beta + C_{k_{T,T+t}}(T+t))^s}\right] \\ &:= \frac{\Gamma(x+r+1)\Gamma(s-1)}{\left(\sum_{j=1}^{x} B_{k_j} + B_{k_T} + \alpha\right)^{x+r+1}} S_{k_{T,T+t}}^*.\end{aligned}$$

By combining everything, we obtain the expression for the conditional expectation reported in Appendix A.3. In Online Appendix EC.5, we show the relationship of the conditional expectation and the DECT.

Appendix D. Additional Model Results for DERT/DECT

In this Appendix, we report detailed results for identifying future top customers using DERT (in the case of the standard Pareto/NBD model) and DECT (in the case of the extended Pareto/NBD model). The prediction performance is very similar to the performance when using conditional expected transactions, because those expressions are closely related. We elaborate on the relationship between these two expressions in detail in Online Appendix EC.5.

Table D.1 shows the predictive performance in identifying the second-tier future customers over the maximum length of the holdout period using DERT and DECT. We observe that the extended Pareto/NBD model outperforms the standard Pareto/NBD model for the first and third data sets. For the second data set, the performance of both models is identical. In the case of the first data set, the extended Pareto/NBD model improves the prediction accuracy by more than 40%.

Appendix E. Specifications for Runtime Comparison

In this Appendix, we provide additional specifications for the runtime comparison presented in Table 20. It shows run times for the the different models applied in the paper for the multichannel catalog merchant (Section 4.1). The data set contains 1,402 customers who had a total of 2,929 transactions over a period of 7.7 years. We fit the models for an estimation period of 1 year and predict during the holdout period of 6.7 years. We report the run times for fitting the model (i.e., likelihood optimization or MCMC sampling).

All timings were measured using a machine featuring 32-CPU cores (Intel Xeon E5-2680 v4 @ 2.40 GHz) and 256 GB of RAM. The relatively large size of the machine was selected because of the high memory requirements of the GPPM, which failed to predict over the full timespan of 6.7 years with less than 256 GB of RAM. All estimations used R (R Core Team 2018) Version 3.5.2.

MCMC settings (if applicable):

- GPPM analogously Dew and Ansari (2018): total iteration, 4,000; initial burn-in, 2,000; chains, 1
- Pareto/GGG analogously Platzer and Reutterer (2016): total iteration, 8,000; initial burn-in, 2,000; chains, 4

Table D.1. Predictive Performance in Identifying the Second-Tier Customers Using DERT/DECT

Second-tier customers		Extended Pareto/NBD	Standard Pareto/NBD
Multichannel catalog merchant	High, correctly classified (%)	59.59	40.86
	Low, correctly classified (%)	81.33	71.53
	Overall correctly classified (%)	74.47	61.84
	High, incorrectly classified (%)	18.67	28.47
	Low, incorrectly classified (%)	40.41	59.14
	Overall incorrectly classified (%)	25.54	38.16
Electronic retailer	High, correctly classified (%)	47.40	47.40
	Low, correctly classified (%)	86.13	86.13
	Overall correctly classified (%)	78.05	78.05
	High, incorrectly classified (%)	13.87	13.87
	Low, incorrectly classified (%)	52.60	52.60
	Overall incorrectly classified (%)	21.95	21.95
Sporting goods retailer	High, correctly classified (%)	55.77	50.77
	Low, correctly classified (%)	85.78	84.22
	Overall correctly classified (%)	78.34	76.10
	High, incorrectly classified (%)	14.22	15.78
	Low, incorrectly classified (%)	44.23	49.23
	Overall incorrectly classified (%)	21.47	23.90

References

Abe M (2009) Counting your customers one by one: A hierarchical Bayes extension to the Pareto/NBD model. *Marketing Sci.* 28(3): 541–553.

Allenby GM, Leone RP, Jen L (1999) A dynamic model of purchase timing with application to direct marketing. *J. Amer. Statist. Assoc.* 94(446):365–374.

Ascarza E, Iyengar R, Schleicher M (2016) The perils of proactive churn prevention using plan recommendations: Evidence from a field experiment. *J. Marketing Res.* 53(1):46–60.

Bachmann P, Naef J, Schilter P, Kuebler N, Meierer M (2020) CLVTools: Tools for customer lifetime value estimation. Accessed September 17, 2020, https://CRAN.R-project.org/package=CLVTools.

Braun M, Schweidel DA, Stein E (2015) Transaction attributes and customer valuation. *J. Marketing Res.* 52(6):848–864.

Colombo R, Jiang W (1999) A stochastic RFM model. *J. Interactive Marketing* 13(3):2–12.

Dew R, Ansari A (2018) Bayesian nonparametric customer base analysis with model-based visualizations. *Marketing Sci.* 37(2): 216–235.

Ebbes P, Papies D, van Heerde HJ (2011) The sense and non-sense of holdout sample validation in the presence of endogeneity. *Marketing Sci.* 30(6):1115–1122.

Ehrenberg AS (1959) The pattern of consumer purchases. *J. Royal Statist. Soc. Ser. C. Appl. Statist.* 8(March):26–41.

Ehrenberg AS (1988) *Repeat-Buying: Facts, Theory, and Applications*, 2nd ed. (Charles Griffin and Company, London).

Fader PS (2012) *Customer Centricity: Focus on the Right Customers for Strategic Advantage* (Wharton Digital Press, Philadelphia).

Fader PS, Hardie BG (2005) A note on deriving the Pareto/NBD model and related expressions. Accessed September 17, 2020, http://www.brucehardie.com/notes/009/pareto_nbd_derivations_2005-11-05.pdf.

Fader PS, Hardie BG (2007) Incorporating time-invariant covariates into the Pareto/NBD and BG/NBD models. Accessed, http://www.brucehardie.com/notes/019/time_invariant_covariates.pdf.

Fader PS, Hardie BG, Lee K (2005a) Counting your customers the easy way: An alternative to the Pareto/NBD model. *Marketing Sci.* 24(2):275–284.

Fader PS, Hardie BG, Lee K (2005b) RFM and CLV: Using iso-value curves for customer base analysis. *J. Marketing Res.* 42(4):415–430.

Fader PS, Hardie BGS, Shang J (2010) Customer-base analysis in a discrete-time noncontractual setting. *Marketing Sci.* 29(6): 1086–1108.

Glady N, Lemmens A, Croux C (2015) Unveiling the relationship between the transaction timing, spending and dropout behavior of customers. *Internat. J. Res. Marketing* 32(1):78–93.

Gopalakrishnan A, Bradlow ET, Fader PS (2017) A cross-cohort changepoint model for customer-base analysis. *Marketing Sci.* 36(2):195–213.

Gupta S (1991) Stochastic models of interpurchase time with time-dependent covariates. *J. Marketing Res.* 28(1):1–15.

Hanssens DM, Pauwels KH (2016) Demonstrating the value of marketing. *J. Marketing* 80(6):173–190.

Harman DM (2016) Stochastic process customer lifetime value models with time-varying covariates. Dissertation, University of Iowa, Iowa City.

Hruschka H (2010) Considering endogeneity for optimal catalog allocation in direct marketing. *Eur. J. Oper. Res.* 206(1):239–247.

Jerath K, Fader PS, Hardie BGS (2011) New perspectives on customer death using a generalization of the Pareto/NBD model. *Marketing Sci.* 30(5):866–880.

Jeuland AP, Bass FM, Wright GP (1980) A multibrand stochastic model compounding heterogeneous Erlang timing and multinomial choice processes. *Oper. Res.* 28(2):255–277.

Knox G, van Oest R (2014) Customer complaints and recovery effectiveness: A customer base approach. *J. Marketing* 78(5):42–57.

Kuhn HW, Tucker AW (1951) Nonlinear programming. Neyman J, ed. *2nd Berkeley Sympos. Math. Statist. Probab.y* (University of California Press, Berkeley), 481–492.

Malthouse EC (2009) The results from the lifetime value and customer equity modeling competition. *J. Interactive Marketing* 23(3): 272–275.

Marketing EDGE (2013) Specialty multi-channel catalog company. Accessed September 17, 2020, http://www.marketingedge.org/marketing-programs/data-set-library.

McCarthy D, Fader P (2018) Customer-based corporate valuation for publicly traded non-contractual firms. *J. Marketing Res.* 55(5): 617–635.

Neslin SA, Rhoads EE (2009) A model and empirical analysis of patient compliance and persistence in pharmaceuticals. Working paper, Dartmouth College, Dartmath, MA.

Netzer O, Lattin JM, Srinivasan V (2008) A hidden Markov model of customer relationship dynamics. *Marketing Sci.* 27(2):185–204.

Ni J, Neslin AS, Sun B (2012) The ISMS durable goods data sets. *Marketing Sci.* 31(6):1008–1013.

Park S, Gupta S (2012) Handling endogenous regressors by joint estimation using copulas. *Marketing Sci.* 31(4):567–586.

Petrin A, Train K (2010) A control function approach to endogeneity in consumer choice models. *J. Marketing Res.* 47(1):3–13.

Platzer M (2018) BTYDplus: Probabilistic models for assessing and predicting your customer base. Accessed September 17, 2020, https://github.com/mplatzer/BTYDplus#readme.

Platzer M, Reutterer T (2016) Ticking away the moments: Timing regularity helps to better predict customer activity. *Marketing Sci.* 35(5):779–799.

R Core Team (2018) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna).

Reinartz WJ, Kumar V (2000) On the profitability of long-life customers in a noncontractual setting: An empirical investigation and implications for marketing. *J. Marketing* 64(4):17–35.

Reinartz WJ, Kumar V (2003) The impact of customer relationship characteristics on profitable lifetime duration. *J. Marketing* 67(1): 77–99.

Rivers D, Vuong Q (1988) Limited information estimators and exogeneity tests for simultaneous probit models. *J. Econometrics* 39(3):347–366.

Ross SM (2014) *Introduction to Probability Models*, 11th ed. (Elsevier, Amsterdam).

Schmittlein DC, Peterson RA (1994) Customer base analysis: An industrial purchase process application. *Marketing Sci.* 13(1):41–67.

Schmittlein DC, Bemmaor AC, Morrison DG (1985) Why does the NBD model work? Robustness in representing product purchases, brand purchases and imperfectly recorded purchases. *Marketing Sci.* 4(3):255–266.

Schmittlein DC, Morrison DG, Colombo R (1987) Counting your customers: Who are they and what will they do next? *Management Sci.* 33(1):1–24.

Schroeder N, Hruschka H (2016) Investigating the effects of mailing variables and endogeneity on mailing decisions. *Eur. J. Oper. Res.* 250(2):579–589.

Schweidel DA, Knox G (2013) Incorporating direct marketing activity into latent attrition models. *Marketing Sci.* 32(3):471–487.

Shugan SM (2004) Endogeneity in marketing decision models. *Marketing Sci.* 23(1):1–3.

Singh SS, Jain DC (2013) Measuring Customer Lifetime Value: Models and Analysis. INSEAD Working Paper No. 2013/27/MKT, http://www.ssrn.com/abstract=2214860.

Singh SS, Borle S, Jain DC (2009) A generalized framework for estimating customer lifetime value when customer lifetimes are not observed. *Quant. Marketing Econom.* 7(2):181–205.

Van Oest R, Knox G (2011) Extending the BG/NBD: A simple model of purchases and complaints. *Internat. J. Res. Marketing* 28(1):30–37.

Wooldridge JM (2015) Control function methods in applied econometrics. *J. Human Resources* 50(2):420–445.

Wübben M, Wangenheim F (2008) Instant customer base analysis: Managerial heuristics often "get it right". *J. Marketing* 72(3):82–93.

Wünderlich R (2015) Hierarchical Bayesian models for predicting purchase behavior in noncontractual settings based on individual and cross-sectional purchase patterns. Dissertation, Technische Universität München, Germany.

Xia F, Chatterjee R, May JH (2019) Using conditional restricted Boltzmann machines to model complex consumer shopping patterns. *Marketing Sci.* 38(4):711–727.

Zhang Y, Bradlow ET, Small DS (2015) Predicting customer value using clumpiness: From RFM to RFMC. *Marketing Sci.* 34(2): 195–208.